

Universitat Politècnica de Catalunya (UPC)

Facultat de Matemàtiques y Estadística

Tesi de màster

**EL SISTEMA ALIMENTARIO GLOBAL: PONDERACIÓN CUANTITATIVA DE
LAS VARIABLES DEL MODELO EN EL ENTORNO DE CATALUÑA**

Francisca Morey Cortès

Ponent de la tesi: Ernest Pons

Tutora de la tesi: Yvonne Colomer

A mi familia y amigos, por su apoyo incondicional;

A la Fundación Triptolemos por darme la oportunidad de introducirme en el proyecto;

En concreto a Ramón Clotet y Eusebi Jarauta, catedrático en bioquímica y catedrático en
matemática aplicada, por sus consejos e ideas;

A Ernest Pons, como ponente, y Yvonne Colomer, como tutora, por sus
aportaciones, siempre constructivas;

A Mireia Fernández, profesora de la UPC e investigadora del IN3, por su
desinteresada y valiosa implicación;

Y, en definitiva, a todas las personas que han contribuido a que
la elaboración de este trabajo haya sido posible;

Gracias

Índice

Resumen y palabras clave.....	5
1 Introducción.....	6
1.1 Objetivos.....	6
1.2 Impacto de la investigación.....	7
1.3 Hipótesis y resultados esperados.....	7
1.4 Estructura del Trabajo Final de Máster.....	8
2 Marco teórico: Antecedentes.....	9
2.1 El Sistema Alimentario Global: I – Definición de un espacio.....	10
2.2 El Sistema Alimentario Global: II – Aproximación cuantitativa al espacio agroalimentario de la Europa mediterránea.....	14
3 Metodología.....	19
3.1 La encuesta.....	19
3.1.1 Identificación del problema.....	19
3.1.2 Determinación del diseño de investigación.....	19
3.1.3 Selección de la muestra.....	20
3.1.4 Técnica de muestreo utilizado: aleatorio estratificado.....	21
3.1.5 Diseño del cuestionario.....	22
3.1.6 Prueba piloto o pre-test.....	23
3.2 Técnicas estadísticas utilizadas para la realización de los análisis posteriores.....	24
3.2.1 Técnica estadística de muestreo: ponderaciones.....	24
3.2.2 Técnica estadística sobre la distribución: bootstrapping.....	24
3.2.3 Técnica estadística de imputación de valores perdidos: Algoritmo <i>Multiple Imputation by Chained Equations (MICE)</i>	25

3.2.4	Técnica estadística para el <i>clustering</i> : distancia Gower, método de ward, test Silhouette y algoritmo Partición Alrededor de Medoids (PAM).....	26
4	Depuración, estructuración y presentación de los datos.....	28
4.1	Depuración, imputación de valores perdidos, agrupación y codificación de los datos.....	28
4.2	Distribución de las respuestas obtenidas versus las esperadas y aplicación de la ponderación.....	29
5	Resultados.....	32
5.1	Estadísticos descriptivos univariantes.....	32
5.1.1	Variables sociodemográficas.....	32
5.1.2	Variables de valoración de la importancia de los cuatro ejes: disponibilidad alimentaria, economía, políticas públicas y conocimiento/saber del sistema alimentario.....	34
5.1.3	Variables sobre práctica y aspiración en el sistema alimentario.....	37
5.2	Análisis estadístico bivalente entre los 4 ejes y las variables sociodemográficas Sexo, Edad y Provincia.....	38
5.3	Análisis de clúster.....	41
5.3.1	Clúster jerárquico.....	41
5.3.2	Clúster no jerárquico.....	42
5.4	Valores de los ratio de posición para España con la ponderación de las medias extraídas de la encuesta para cada variable.....	46
6	Discusión.....	47
7	Conclusiones.....	51
	Bibliografía.....	53
	Anexo I Cuestionario.....	54
	Anexo II Código de R.....	57

RESUMEN

El presente trabajo constituye la tercera parte de “El Sistema Alimentario Global”. El objetivo principal es proponer la encuesta como herramienta para optimizar el modelo matemático que se llevó a cabo en las dos fases anteriores y que se basaba en una red que configuraba el espacio alimentario a través de 4 ejes diferentes: Disponibilidad alimentaria, Economía, Políticas y Cultura, los cuales se les daba una ponderación igual a cada uno. Con el perímetro definido a Cataluña se pretende, a partir de la encuesta, saber qué importancia le dan los catalanes a estos ejes y así poder aplicar una ponderación más afinada y representativa de la realidad. Para ello se lleva a cabo el diseño del cuestionario, la recopilación de los datos, la aplicación de técnicas estadísticas de muestreo: estratificación por cuotas y ponderación de los estratos, técnicas de depuración de encuestas, así como técnicas de “profiling” para diferenciar grupos entre las variables del sistema alimentario y las características sociodemográficas que los determinan.

Palabras clave: sistema alimentario global, disponibilidad alimentaria, economía, política, cultura, encuesta, Cataluña.

ABSTRACT

The present paper constitutes the third part of “The Global Food System”. The main objective is to propose the survey as a technique to optimize the mathematical model that was carried out in the two previous phases and that was based on a network that configured the food space through 4 different axes: Availability, Economy, Policies and Culture, which were given an equal weight to each. With a defined perimeter to Catalonia, it is intended, from the survey, to know what importance the catalans give to these axes and thus, to be able to apply a more refined and representative weighting to reality. For this purpose, the design of the questionnaire, the collection of the data, the application of statistical sampling techniques are carried out: stratification by ranges and weighting of the strata, debugging techniques, as well as profiling techniques to differentiate groups between the variables of the food system and the sociodemographic characteristics that determine them.

Keywords: food system global, food availability, economy, politics, culture, survey, Catalonia.

1. Introducción

Es indudable que, des de diversas perspectivas, interesa estudiar y caracterizar el sistema alimentario de una región, un país, conjunto de países, ... ya que se trata de un pilar imprescindible para la supervivencia y crecimiento de una sociedad. Esto se traduce a querer entender su funcionamiento, su metodología, sus puntos fuertes así como sus carencias. Además, para los estados políticos (entendidos como mecanismos de regulación de la sociedad y responsables en la creación de políticas públicas) tener investigaciones sobre el tema y fomentadas en datos, análisis e investigaciones rigurosas, da a la toma de decisiones una mayor credibilidad y coherencia. Lo demuestra el hecho de que se trate de un proyecto reconocido y priorizado por el Ministerio de Agricultura, Pesca y Alimentación. De la misma manera lo es estudiarlo a nivel global: tener una visión del papel que juega el sistema alimentario teniendo en cuenta todas las áreas que influyen en él, lo que le añade un poder substancial y pone de manifiesto el interés que tiene.

Por eso, este trabajo trata de seguir con la investigación del modelo cuantitativo sobre el sistema alimentario global que lleva a cabo la Fundación Triptolemos. Concretamente, se trata de la tercera fase del proyecto, donde las dos primeras se basaron en:

1. Definición del espacio. Es decir, desarrollar la primera visión del modelo, exponer la metodología matemática que se quería implementar y diseñar los ejes a tener en cuenta como influyentes en el sistema alimentario global: economía, saber/factores culturales, política y disponibilidad alimentaria.
2. Aproximación cuantitativa al espacio alimentario de la Europa Mediterránea. Esta segunda fase tuvo como objetivo el implementar el modelo creado en la primera en cuatro países diferentes: España, Francia, Grecia e Italia. Para ello se tuvo que definir las variables que explicaban mejor los ejes del sistema alimentario. Uno de los resultados más relevantes fue la creación del Índice Triptolemos que permite de manera cuantitativa la comparación entre sistema alimentarios en diferentes países y territorios.

1.1. Objetivos

Esta tercera fase tiene como objetivo el diseño de una metodología para saber qué ponderación se

le tiene que dar a cada eje, ya que se entiende que no es lo mismo la disponibilidad alimentaria o la economía que se tiene de los alimentos que las políticas del Estado en referencia a la alimentación o el saber y/o factores culturales que tiene una sociedad sobre el tema. En la segunda fase, cuando se implementó el modelo a los cuatro países, se aplicó la misma ponderación a cada uno, 0,25 puntos, haciendo un total de 1, lo que se consideró un punto débil a mejorar. De manera que este trabajo pretende:

1. Proponer una herramienta para conseguir saber la ponderación de los ejes.
2. Aplicarla para obtener los datos.
3. Analizar los datos obtenidos.
4. Hacer un “profiling” a partir de los cuatro ejes a analizar para saber que características sociodemográfico determina la importancia de cada eje.
5. Elaborar los análisis y hacer una comparativa de los resultados sin y con la ponderación, la cual la última va a reflejar mucho mejor la realidad. Concretamente observar como cambia el Índice Triptolemos creado por la Fundación y que tiene como objetivo evaluar el sistema alimentario del territorio estudiado.

1.2. Impacto de la investigación

Por tanto, la relevancia y el interés de este trabajo yace en el hecho de seguir modelando el sistema alimentario global de una manera cuantitativa y además perfilando la rigurosidad de su metodología y su aplicación, dando como resultado una herramienta muy útil para decisiones que se quieran tomar sobre el tema, ya sean políticas como empresariales. De hecho, el artículo que se realizó en la primera fase del proyecto, se publicó en el Revista Española de Estudios Agrosociales y Pesqueros y se trata de un proyecto con financiación del Ministerio de Industria, Turismo y Comercio. Además, se le añade que ha sido reconocido por la UNESCO con el otorgamiento de la Cátedra “*Ciencia e innovación para el desarrollo sostenible: producción y seguridad alimentaria mundiales*” (Triptolemos, 2018).

1.3. Hipótesis y resultados esperados

Como herramienta elegida para plasmar de manera idónea la ponderación de los ejes, ha sido la encuesta. Se ha diseñado una encuesta y como marco territorial elegido ha sido Cataluña. En consecuencia, las hipótesis que se plantean y los resultados esperados están basados en este marco. Concretamente, son los siguientes:

1. Una importancia mayor en los ejes de Disponibilidad alimentaria y Economía, justificado a partir del concepto de seguridad alimentaria, es decir, como ejes imprescindibles para la obtención de alimentos.
2. Una ponderación menor en los ejes de Saber y Políticas, entendido como aspectos que influyen en la alimentación de una manera indirecta y por tanto, no percibida como indispensable para la sociedad.
3. Diferencias en la ponderación de los ejes y las variables sociodemográficas. El sexo, la edad o el territorio son factores que influyen en la valoración de los diferentes ejes.
4. Cataluña, como país con una tendencia alimentaria de Occidente, le da una mayor importancia a la soberanía alimentaria (detenimiento al sabor nutricional de los alimentos) que a la globalización alimentaria (mayor disponibilidad de alimentos en la sociedad) (Stuart, M., 2016).

Por tanto, los resultados esperados a partir de este trabajo es ver si se cumplen o no las hipótesis anteriores y poder modelar las diferencias de los ejes de la población de Cataluña a partir de las variables del cuestionario.

1.4. Estructura del Trabajo Final de Máster

El trabajo se divide en 4 grandes capítulos:

En el capítulo 2. *Marco Teórico: Antecedentes* se explican de manera resumida las dos anteriores fases del proyecto a través de las publicaciones hechas: la primera *“El Sistema Alimentario Global: I – Definición de un espacio”* y la segunda *“El sistema alimentario global: II – Aproximación cuantitativa al espacio agroalimentario de la Europa mediterránea.”*

Seguidamente, en el capítulo 3. *Metodología* se hace referencia a todas las técnicas que se han utilizado en el trabajo, tanto la parte cualitativa con la creación del cuestionario como la parte

cuantitativa, más amplia y que engloba la codificación, la imputación de valores perdidos, la aplicación de ponderaciones y las librerías utilizadas, técnica de bootstrapping para hacer frente a la no normalidad de los datos, la técnica utilizada para el análisis de clúster y diferentes aspectos técnicos y metodológicos que se han tenido en cuenta para alcanzar los resultados de manera rigurosa.

El capítulo 4. *Presentación de los datos* hace referencia a los pasos previos que se han hecho en la base de datos para poder luego hacer los análisis. Concretamente se trata observar cómo se distribuyen las respuestas obtenidas versus las esperadas, la aplicación de las ponderaciones, la imputación de valores perdidos, la depuración de los datos, la codificación y la agrupación de variables.

En el capítulo 5. *Resultados* se analizan los datos obtenidos en la encuesta. Por un lado, se han realizado análisis univariantes para entender los resultados globales de Cataluña. Por otro lado, se contrastan estos análisis con las variables sociodemográficas para observar los diferentes perfiles. Un tercer subapartado, es el análisis de clúster para ver los diferentes grupos que encontramos en Cataluña en referencia a los ejes sobre la alimentación y así poder establecer perfiles. Y finalmente, un cuarto subapartado donde se aplican las ponderaciones extraídas de la encuesta a los datos que se analizaron en la fase dos del proyecto sobre los países mediterráneos europeos.

En el capítulo 6. *Discusión* se pone énfasis a la interpretación de los resultados obtenidos en el anterior apartado. Conjuntamente con literatura sobre el sistema alimentario, se propone un discurso crítico teniendo en cuenta todos los aspectos del trabajo, las investigaciones anteriores y las hipótesis establecidas en esta tercera fase.

Y finalmente, el capítulo 7. *Conclusiones* encontramos, a través de un argumento lógico derivado de la evaluación de las premisas y el pensamiento crítico, de argumentos y afirmaciones, las conclusiones a que se han llegado después de la realización del trabajo.

2. Marco teórico: antecedentes

En este apartado lo que se pretende es entender la investigación que se ha hecho anteriormente a este trabajo y así comprender de una manera más completa el objetivo, el contexto y la relevancia de éste.

Concretamente, se hará mención a los dos artículos que se publicaron y que tenían como objetivo crear un modelo que representara el sistema alimentario global y la derivación a un caso práctico: I – Definición de un espacio y II – Aproximación cuantitativa al espacio agrolimentario de la Europa mediterránea.

2.1 El Sistema Alimentario Global: I – Definición de un espacio

En este artículo los autores¹ desarrollan un modelo del sistema alimentario global a través de un espacio multidimensional, basado en cuatro ejes básicos con el objetivo que cualquier actividad del sistema alimentario pueda verse representada y situada en al menos uno de ellos. Específicamente son los siguientes:

- **Disponibilidad (D):** consideración de los elementos que permitan asegurar que cuantitativamente haya alimentos suficientes para la alimentación correcta del ser humano. Y para la medición del eje se tienen en cuenta tres dimensiones:
 - Producción primaria (d_1): productos que se obtengan por la vía de la fotosíntesis.
 - Producción de productos elaborados y servicios (d_2): todo producto que ha tenido alguna intervención/transformación por las empresas alimentarias.
 - Nutrición (d_3): toda actividad científica o técnica que contenga conceptos referidos a los mecanismos nutricionales. Concretamente se mide el consumo individual de alimentos.
- **Economía (E).** Este eje vendría definido por las siguientes dimensiones:
 - Economía agroalimentaria total (e_1): concepto entendido como toda actividad económica alimentaria, desde el campo hasta el momento previo a la ingesta del alimento. (Clotet, R., 2013), con excepción de los servicios.
 - Economía del ciudadano (e_2), aspecto que se recoge a través del Producto Interior Bruto (PIB).
 - Economía global entendida como la facilidad de desarrollo económico y social (e_3). Este concepto es considerado a través del coste de la energía total consumida y se expresa

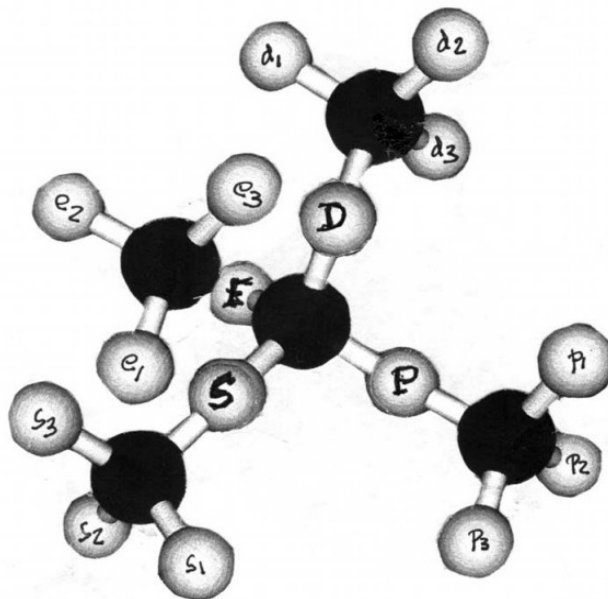
¹ R. Clotet Ballús, Y. Colomer Xena, E. Jarauta y F. Mayor Zaragoza

en euros/habitantes por año.

- **Política (P):** toda actividad que la sociedad, entendida como ente político, genera en torno al sistema alimentario. Concretamente, en los siguientes ámbitos:
 - Derechos humanos (p_1). En concreto se considera el nivel de calidad de vida como indicador global de buen funcionamiento del sistema alimentario.
 - Acciones públicas de los gobiernos sobre el sistema (p_2). Para analizar este concepto se toma como variables el presupuesto oficial del Ministerio de Agricultura de cada Administración.
 - Acciones de ayuda (p_3) entendidas como los esfuerzos de la sociedad civil en contribuir a través de organizaciones a minimizar las situaciones de carencia alimentaria.
- **Saber (S).** Esta esfera se plantea a través de tres áreas diferentes:
 - Conocimiento (s_1): se refiere a la preocupación/inquietud por los alimentos y la nutrición de los ciudadanos. Como variable para representarlo se ha tomado el número de publicaciones clasificadas en el apartado global de Agricultura y de Nutrición.
 - Comportamiento social (s_2): acciones alimentarias explicadas a través de la historia, sociología, antropología, ...Este concepto se representa a través de la variable que cuantifica la evolución en el tiempo del consumo de carne de todo tipo en la dieta.
 - Cultura (s_3): entendida como reflexión tanto intelectual como de placer relacionada con los alimentos y que se representa con la variable de número de publicaciones en el apartado de “*Arte culinario*” de la clasificación CDU (Clasificación Decimal Universal).

A partir de estos cuatro ejes y sus subdivisiones se consigue disponer de doce componentes que se relacionan entre ellos en combinaciones binarias, creando una red [figura 1].

Figura 1. RED POLIÉDRICA ESPACIAL DEL SISTEMA ALIMENTARIO GLOBAL



Fuente: Fundación Triptolemos

Esta red configura un espacio representado como un tetraedro regular, donde cada uno representa el núcleo inicial espacial: disponibilidad (*D*), economía (*E*), política (*P*) y saber (*S*). Este modelo tiene como objetivo representar el sistema alimentario global a través de establecer valores cuantitativos a partir de variables para cada uno de los subejos, así como de medir la distancia del centro a éstos dependiendo de la participación de cada una de ellas.

Este modelo también está planteado con la visión de poder crecer sin limitaciones: el aumento de precisión vendrá definido por el incremento de subejos representados como otro tetraedro o si es preciso, otra figura geométrica.

La metodología utilizada para llevar acabo el análisis del sistema alimentario a través de una red tetraédrica, está basada en lo que se denomina análisis composicional, es decir, como un sistema dinámico caracterizado por variables cuantitativas, de valores positivos que evolucionan en el tiempo. De esta manera, se pretende realizar proyecciones de cómo puede ser en el futuro la evolución de los grupos de población.

Un sistema dinámico que está subdividido en m grupos puede describirse como una función

vectorial:

$$\vec{F}: I \subseteq \mathbb{R} \rightarrow \mathbb{R}^m; \quad \vec{F}(t) = (F_1(t), F_2(t), \dots, F_m(t)), t \in I \quad (1)$$

Los componentes de esta función tienen unidades denominadas *valores en masa*, tales como número de personas, unidades monetarias, número de barriles, ... y configuran una matriz en que se definen las proporciones o partes como el cociente de cada una de los componentes:

$$M(t) = F_1(t) + F_2(t) + \dots + F_m(t), t \in I; \quad x_k(t) = \frac{F_k(t)}{M(t)}, k = 1, 2, \dots, m \quad (2)$$

Y se obtiene una función simplicial: una recta real en el denominado *símplex* de m partes:

$$\vec{x}: I \subseteq \mathbb{R} \rightarrow S^m; \quad \vec{x}(t) = (x_1(t), x_2(t), \dots, x_m(t)), t \in I; \quad \sum_{k=1}^m x_k(t) = 1 \quad (3)$$

En el caso del sistema alimentario, se plantean unas variables escogidas de fuentes estadísticas fiables (organismos internacionales y públicos) y expresadas en unidades homogéneas, que permiten un análisis cuantitativo. Concretamente, las variables elegidas para cada subjeje son las siguientes [tabla 1]:

Tabla 1. CONCEPTOS DE CADA SUBEJE Y VARIABLE DE MEDIDA ESCOGIDA EN CADA CASO

Eje	Subeje	Concepto	Variable propuesta o unidades de expresión
D	D1	Producción primaria	Kg/habitante año
	D2	Producción transformación y elaborados	Kg/habitante año
	D3	Nutrición (*)	Kg/habitante año
E	E1	Valor económico de producción primaria y de productos elaborados	€/habitante año
	E2	Producto Interior Bruto (PIB)	€/habitante año
	E3	Valor económico energía total consumida	€/habitante año
P	P1	Derechos humanos. Exclusión social	Personas en NO riesgo de exclusión social/ Habitantes año
	P2	Acciones públicas sobre el sistema agroalimentario	Presupuesto oficial estructura agroalimentaria en relación al presupuesto total €/habitante año
	P3	Acciones de ayuda a necesidades sociales de alimentos (sociedad civil)	Kg. Alimentos ayuda/habitante año
S	S1	Conocimiento general del mundo agroalimentario	Publicaciones generales sobre el mundo agroalimentario/habitante año
	S2	Comportamiento en la dieta	Consumo de carnes en kg/habitante año
	S3	Cultura con base agroalimentaria	Publicaciones sobre arte culinario/habitante año

Fuente: Fundación Triptomelos

Debido a la estructura algebraica y geométrica del símplex que no permite aplicar los métodos estadísticos estándares, se hace una transformación logcociente isométrica (*ilr*). De esta manera se puede efectuar el ajuste de un modelo de regresión lineal a una ecuación diferencial lineal de primer orden. Esta ecuación corresponde al sistema de ecuaciones diferenciales:

$$\frac{d\vec{u}(t)}{dt} = A\vec{u}(t) + \vec{b} \quad (4)$$

Esta estimación presenta ciertos problemas de estabilidad numérica, de manera que se trabaja con la forma integral del sistema anterior:

$$\frac{du_j(t)}{dt} = a_1^j u_1(t) + a_2^j u_2(t) + \dots + a_{m-1}^j u_{m-1}(t) + b_j, \quad j = 1, 2, \dots, m-1 \quad (5)$$

Con la regresión lineal se estiman los coeficientes del sistema y los términos independientes.

$$\begin{aligned} u_j(t) - u_j(0) &= a_1^j U_1(t) + a_2^j U_2(t) + \dots + a_{m-1}^j U_{m-1}(t) + b_j t, \\ U_j(t) &= \int_0^t u_j(z) dz, \quad j = 1, 2, \dots, m-1 \end{aligned} \quad (6)$$

En definitiva, con la aplicación de los datos al modelo se obtiene la visualización de la prospectiva de las variables analizadas, conocer las relaciones en el tiempo y en la geografía de sus parámetros y realizar estrategias prospectivas sobre el mismo. A partir de la presentación de este modelo y su metodología, se procedió a realizar la segunda publicación del proyecto.

2.2 El sistema alimentario global: II – Aproximación cuantitativa al espacio agroalimentario de la Europa mediterránea.

En esta segunda parte de la investigación, a partir del modelo planteado anteriormente, se procedió a realizar con detalle y caracterizar la evolución en el tiempo de los factores específicos de cada uno de los ejes considerados en el diseño fractal del modelo en un área geográfica concreta: la Europa mediterránea. Específicamente se tuvieron en cuenta los países de referencia: España, Francia, Grecia e Italia. Las razones de su elección fueron las siguientes:

- Las condiciones geográficas de su producción agroalimentaria

- Las costumbres alimentarias como resultado histórico del anterior punto
- Los aspectos de mercado, economía, legislación general y políticas fruto de la pertenencia a la Unión Europea (UE)

Con la elección de los países, de las variables que se han detallado en la primera parte de la investigación [tabla 1] y la aplicación de los datos en cada uno de ellos, se procede a:

1. Ilustrar y desarrollar el modelo cuantitativo de análisis comparativo
2. Definir un factor de posición del sistema alimentario de cada país
3. Definir un índice de calidad global del sistema alimentario de cada país en un determinado entorno

Para llevar a cabo la comparativa de los cuatro países a través de las variables de los cuatro ejes, se estandarizan para hacer referencia a valores relativos (por habitante). Para su tratamiento, la metodología utilizada se fundamenta en el **Análisis de Datos Composicionales** (Aitchison, 1986) y que se basa en (Jarauta, E., 2018):

1. Disponer en una tabla los valores para la serie de años de la que se dispone de datos.
2. Calcular los valores relativos para cada habitante mediante el cociente entre el correspondiente total y el número de habitantes de cada país.
3. Calcular las proporciones, es decir, lo que aporta cada país en el total (cociente de cada dato y la suma de todos los del mismo año).
4. Cuantificar la posición relativa de cada país en relación al conjunto, representada por su media geométrica a partir del logaritmo neperiano (cociente entre la proporción de cada país y la media geométrica de las proporciones de la serie). Este cálculo se conoce como *transformación logcociente centrada (CLR)* y su implicación es el conseguir lo que se designa *Posición Relativa (PR)*, es decir, que los valores positivos están situados por encima del valor promedio y los valores negativos por debajo.
5. Aplicar el modelo a partir de ecuaciones diferenciales lineales composicionales. La estimación es de un período adicional del 1/3.
6. Elaborar una tabla reflejando los valores RP iniciales, actuales y futuros.

7. Aplicar este procedimiento para cada una de las variables.

A partir de la metodología explicada anteriormente, se obtienen los siguiente resultados para cada variables y cada país [tabla 3]:

Tabla 2. VALORES DE LOS RATIO DE POSICIÓN RP Y VALORES MEDIOS (POR CADA TIEMPO) PARA CADA UNO DE LOS PAÍSES Y CADA UNA DE LAS VARIABLES

España	Inicial	Actual	Futuro	Media
D1	-0.085	0.041	0.048	0.002
D2	-0.122	-0.019	-0.022	-0.055
D3	-0.086	-0.080	-0.075	-0.080
E1	-0.066	0.155	0.233	0.107
E2	-0.074	-0.054	-0.059	-0.062
E3	0.054	0.155	0.263	0.157
P1	0.015	-0.007	-0.030	-0.008
P2	0.005	-0.096	-0.182	-0.091
P3	-0.299	-0.208	-0.138	-0.215
S1	0.315	0.258	0.206	0.260
S2	0.481	0.550	0.591	0.541
S3	0.296	0.217	0.160	0.225
Media	0.036	0.076	0.083	0.065

Francia	Inicial	Actual	Futuro	Media
D1	0.196	0.343	0.390	0.309
D2	0.277	0.441	0.407	0.375
D3	0.039	0.006	-0.005	0.013
E1	0.141	0.070	-0.003	0.070
E2	0.204	0.302	0.369	0.292
E3	0.146	-0.111	-0.385	-0.117
P1	0.073	0.093	0.083	0.083
P2	-0.023	0.083	0.174	0.078
P3	0.429	0.339	0.270	0.346
S1	0.256	0.189	0.127	0.191
S2	0.233	0.141	0.086	0.154
S3	0.080	0.295	0.450	0.275
Media	0.171	0.163	0.164	0.172

Grecia	Inicial	Actual	Futuro	Media
D1	0.111	-0.047	-0.098	-0.012
D2	0.050	-0.098	-0.077	-0.042
D3	0.000	0.042	0.042	0.028
E1	0.083	-0.033	-0.046	0.001
E2	-0.242	-0.364	-0.426	-0.344
E3	-1.036	-0.464	0.144	-0.452
P1	-0.076	-0.063	-0.034	-0.057
P2	0.018	-0.062	-0.130	-0.058
P3	-0.289	-0.213	-0.153	0.218
S1	0.000	0.013	0.025	0.013
S2	-0.567	-0.577	-0.583	-0.576
S3	-0.630	-0.587	-0.556	-0.591
Media	-0.215	-0.204	-0.158	-0.192

Italia	Inicial	Actual	Futuro	Media
D1	-0.245	-0.317	-0.340	-0.301
D2	-0.205	-0.324	-0.308	-0.279
D3	0.046	0.032	0.038	0.039
E1	-0.158	-0.194	-0.184	-0.178
E2	0.111	0.117	0.115	0.114
E3	0.503	0.448	-0.022	0.310
P1	-0.012	-0.023	-0.018	-0.018
P2	0.001	0.075	0.139	0.072
P3	0.159	0.081	0.021	0.087
S1	-0.572	-0.461	-0.358	-0.463
S2	-0.147	-0.115	-0.095	-0.119
S3	0.253	0.075	-0.054	0.091
Media	-0.022	-0.050	-0.089	-0.054

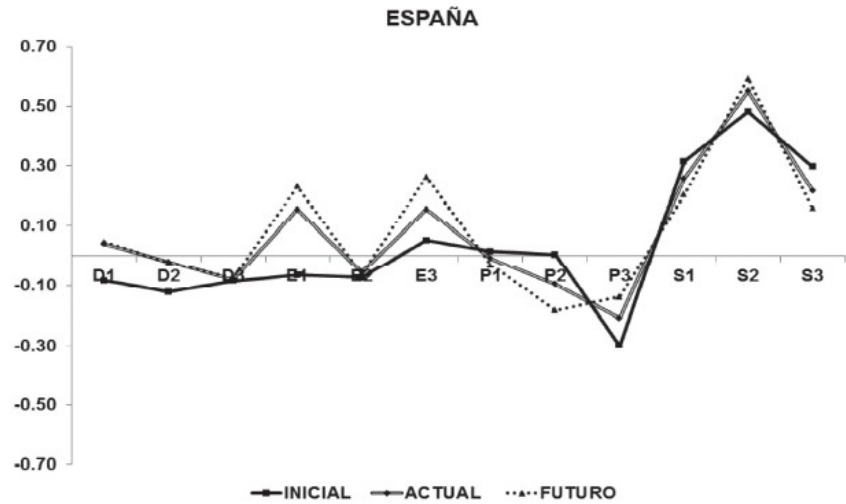
Fuente: FundaciónTryptolemos

Aunque se extrajeron resultados para cada país, en este trabajo nos interesa ver los obtenidos de

manera global para el conjunto de países y también los de España ya que la aplicación hecha en esta investigación es sobre Cataluña.

En el siguiente gráfico podemos ver como se ilustra la evolución de los valores medios de las 12 variables para España [Figura 2]:

Figura 2. ILUSTRACIÓN DE LOS VALORES DE LA RATIO DE POSICIÓN PARA CADA VARIABLE Y EN CADA INSTANTE DE TIEMPO ESTADO (INICIAL, ACTUAL, FUTURO)



Fuente: Fundación Triptolemos

La contribución parcial a través la media aritmética de la RP proporciona el que se denominó en la investigación como **Índice Triptolemos_n (ITRI_n)** y que pone de manifiesto la valoración de cada país en un entorno definido por las variables escogidas. Dicho de otra manera, indica el equilibrio del sistema alimentario de manera global (económico, político y social) que tiene aquel país. Muestra entonces una amplitud de información y una gran posibilidad de correlaciones y comparaciones.

Tabla 3. VALORES DEL ITRI₁₂ PARA LOS CUATRO PAÍSES ESTUDIADOS Y PORCENTAJES DE VALORES NEGATIVOS INDIVIDUALES DE LA RP

	ITRI ₁₂ inicial	ITRI ₁₂ actual	ITRI ₁₂ futuro	ITRI ₁₂ global	Porcentaje de valores negativos de las ratios de posición en el índice global
España	0.036	0.076	0.083	0.065	50.00 %
Francia	0.171	0.163	0.164	0.172	8.33 %
Grecia	-0.215	-0.204	-0.158	-0.192	66.67 %
Italia	-0.022	-0.050	-0.089	-0.054	50.00 %

Fuente: Fundación Triptolemos

Como se muestra en la tabla anterior [tabla 3] se representa para cada uno de los países los

valores del $ITRI_{12}$ en los tiempo inicial, actual y futuro. También se indica el porcentaje de valores negativos correspondiente a los valores medios globales.

Como primera interpretación de los resultados, podemos afirmar Francia como principal país con más potencial y crecimiento equilibrado del sistema alimentario de entre los cuatro países mediterráneos estudiados.

En el caso de España se observa cómo se predicen *“aumentos en la producción primaria y disminución en la transformación y en la accesibilidad alimentaria de la población, coincidentes con la prospectiva de incremento de la población en riesgo de exclusión social (D1, D2, E2, P1)”* (Jarauta, E., 2018). De la misma manera hay una variable que indica un empobrecimiento social en relación con nuestro entorno de países ($D3, E2, P1, P2$ y $P3$), aunque la afinación se mantiene en los temas agroalimentarios y culinarios ($S1, S2, S3$). Concretamente el consumo de carne seguirá incrementando según los datos.

En España en los próximos 5 años habrá un incremento en la actividad económica de producción primaria, tanto en producción como en valor económico y también en el interés cultural por la gastronomía. En cambio, la previsión de la posición relativa en elaboración de transformados será inferior a la media, igual que un mayor coste en energía y una pérdida en competitividad en el entorno, en el PIB, en los fondos de los Estados destinados a políticas agroalimentarias y en la generosidad alimentaria de la sociedad civil.

Y finalmente, comparando España con el $ITRI_{global}$ se observa como está ligeramente por encima de la media entre los países estudiados.

En conclusión, a través de la aplicación de las variables en el modelo y para los cuatro países diferentes se ha podido detectar tendencias nutricionales y por tanto, *“propiciar la toma de decisiones y propuestas que permitan actuar en consecuencia”* (Jarauta, E., 2018). Así el modelo demuestra su gran potencial, cuantificando el sistema alimentario, teniendo en cuenta su globalidad a través de todos los aspectos que tienen una influencia sobre él. Cabe decir, que la precisión de los resultados depende del número de variables que se consideran (cuántas más variables tiene el modelo, más preciso es).

3. Metodología

3.1. La encuesta

Para conseguir el objetivo del trabajo: saber la ponderación de los cuatro ejes (Disponibilidad alimentaria, Economía, Saber y Conocimiento) se ha decidido utilizar la encuesta como técnica de investigación idónea para plasmar esa realidad ya que permite obtener datos de un modo rápido y eficaz. Concretamente, esta decisión se sustenta a través de las cuatro siguientes causas (Anguita, J. C., 2003):

1. La encuesta permite aplicaciones a territorios que mediante técnicas de muestreo adecuadas pueden hacer extensivos los resultados a comunidades enteras.
2. El investigador no es el sujeto de interés, sino la población a la que pertenece.
3. Permite obtener datos muy detallados a lo que se quiere investigar.
4. La información que recoge, al estar estandarizada, proporciona un gran potencial para comparaciones intergrupales.

El proceso del cuestionario se ha llevado a cabo siguiendo etapas que se mostrarán en los siguientes apartados.

3.1.1. Identificación del problema

La definición clara y precisa del objetivo a investigar: la importancia que le da la sociedad catalana a los diferentes ejes de nuestro modelo. Como explica J. Casas Anguita en su artículo *“La encuesta como técnica de investigación. Elaboración de cuestionarios y tratamiento estadístico de los datos”*, el objeto de estudio tiene que estar fomentado y apoyado de información y bibliografía. En nuestro caso, como hemos visto en el anterior capítulo, se ha realizado un marco teórico estableciendo el conocimiento sobre el tema mediante documentación.

3.1.2. Determinación del diseño de investigación

Aquí hay dos puntos de interés. Por un lado, si nos fijamos en el método de investigación para nuestro trabajo se ha elegido el analítico experimental, es decir, que el investigador puede ejercer

control directo sobre las variables para comprobar los efectos del objeto de estudio y, de esta manera, poder determinar cómo es la relación causa-efecto entre los ejes del sistema alimentario y la sociedad catalana.

Por otro lado, si atendemos a la dimensión temporal se trata de una encuesta transversal, es decir, tiene como fin estimar la frecuencia del fenómeno en un momento temporal concreto. Así se podrá establecer diferencias entre distintos grupos que componen la muestra gracias a su muestreo seleccionado aleatoriamente a partir de segmentos.

3.1.3. Selección de la muestra

En este punto se tuvo que decidir qué población se quería estudiar. Como ya se ha dicho anteriormente, se eligió como perímetro de muestreo Cataluña ya que, por razones de tiempo, de costes y de aproximación teórica, era un nivel territorial asequible. Y la franja de edad que se estableció fue de los 15 años o más.

Teniendo en cuenta la siguiente fórmula (7) para calcular el tamaño de muestra cuando se sabe la población, determinamos que la muestra necesaria para que sea representativa de Cataluña es de 385 respuestas, ya que se trata de una población de 6.501.509 habitantes (INE, 2019), un intervalo de confianza del 95% ($k = 1,959964$), una probabilidad de éxito y fracaso del 50% ($p=q=0,5$), con un error del 5% ($e = 0,05$)

$$n = \frac{k^2 \cdot p \cdot q \cdot N}{(e^2 \cdot (N-1)) + k^2 \cdot p \cdot q} \quad (7)$$

En donde, N = tamaño de la población

K = nivel de confianza,

p = probabilidad de éxito, o proporción esperada

q = probabilidad de fracaso

e = precisión (Error máximo admisible en términos de proporción).

3.1.4. Técnica de muestreo utilizado: muestreo aleatorio estratificado

Este tipo de muestreo divide la muestra en subpoblaciones, denominadas estratos, en función de las variables que pueden tener influencia sobre las características que se quiere medir. En nuestro caso al tratarse de un territorio geográfico como objeto de estudio, se ha procedido a hacer los estratos a través de tres variables:

- **Género:** hombre y mujer.
- **Provincia:** Barcelona, Girona, Lleida y Tarragona.
- **Edad:** de 15 a 24 años, de 25 a 34 años, de 35 a 44 años y de 45 o más años.

Primero se procedió a analizar cómo se distribuye la población de Cataluña en los diferentes estratos [Tabla 4]:

Tabla 4. DISTRIBUCIÓN DE LOS HABITANTES DE CATALUÑA POR SEXO, EDAD Y PROVINCIA

Hombres	15-24	25-34	35-44	De 45 o más	TOTAL
Cataluña	396.521	452.630	639.027	1.678.096	3.166.274
Barcelona	290.693	337.661	470.645	1.223.429	2.322.428
Girona	41.677	44.879	63.971	170.713	321.240
Lleida	22.625	25.527	37.112	101.843	187.107
Tarragona	41.526	44.563	67.299	182.111	335.499

Mujeres	15-24	25-34	35-44	De 45 o más	TOTAL
Cataluña	371.214	451.948	616.125	1.895.948	3.335.235
Barcelona	273.599	339.891	459.330	1.413.793	2.486.613
Girona	38.100	44.432	60.746	181.814	325.092
Lleida	20.355	23.617	32.359	105.627	181.958
Tarragona	39.160	44.008	63.690	194.714	341.572

Fuente: elaboración propia a través de datos del INE

El segundo paso es la afijación por las subpoblaciones determinadas de la muestra. La distribución de los 385 encuestados que se necesitan, es la siguiente:

Tabla 5. DISTRIBUCIÓN DE LOS 385 ENCUESTADOS POR SEXO, EDAD Y PROVINCIA

Hombres	16-24	25-34	35-44	De 45 o más	TOTAL
Cataluña	22	28	38	99	187
Barcelona	17	20	28	72	137
Girona	2	3	4	10	19
Lleida	1	2	2	6	11
Tarragona	2	3	4	11	20

Mujeres	16-24	25-34	35-44	De 45 o más	TOTAL
Cataluña	22	27	36	113	198
Barcelona	16	20	27	84	147
Girona	2	3	4	11	20
Lleida	1	1	2	6	10
Tarragona	2	3	4	12	21

Fuente: elaboración propia

3.1.5. Diseño del cuestionario

Al tener claro nuestro objetivo: saber la importancia que mostraba la población catalana de los cuatro ejes del sistema alimentario (disponibilidad alimentaria, economía, política y saber), las preguntas fueron planteadas con la finalidad de obtener esta información [Tabla 6].

Tabla 6. DICcionario DE VARIABLES: CÓDIGO, ETIQUETA, VALORES, TIPO I NATURALEZA DE LA PREGUNTA

Código de variable	Etiqueta	Valores de la variables	Tipo de pregunta	Naturaleza
P1	Sexo	1 Hombre, 2 Mujer	Cerrada	Sociodemográfica
P2	Fecha de nacimiento	dd-mm-yyyy	Semiabierta	Sociodemográfica
P3	Comarca de residencia	Desplegable de las 41 comarcas de Cataluña	Múltiple excluyente	Sociodemográfica
P4	¿Cuál es su nivel de estudios?	1 Educación infantil, 2 Educación primaria, 3 Educación secundaria, 4 Educación postsecundaria, 5 Educación universitaria, 6 Educación no formal	Múltiple excluyente	Sociodemográfica
P5	Si tuviera que elegir una clase social, ¿a cuál diría que pertenece?	1 Baja, 2 Media-baja, 3 Media, 4 Media-alta, 5 Alta	Múltiple excluyente	Sociodemográfica
P6	En qué tipo de alimentación se encuentra más cómodo	1 Predominancia de vegetales, 2 Sin preferencias, 3 Predominancia de carne y derivados	Múltiple excluyente	Práctica alimentaria
P7	Valoración de importancia: Tener disponibilidad de acceso a alimentos nutricionalmente adecuados a la dieta:	1 Poco importante, 2, 3, 4, 5, 6, 7, 8, 9, 10 Muy importante	Pregunta de estimación (ítem de Likert)	Opinión
P8	Valoración de importancia: Tener el nivel económico para comprar los alimentos necesarios	1 Poco importante, 2, 3, 4, 5, 6, 7, 8, 9, 10 Muy importante	Pregunta de estimación (ítem de Likert)	Opinión
P9	Valoración de importancia: Las políticas públicas para favorecer la producción de alimentos y la seguridad alimentaria:	1 Poco importante, 2, 3, 4, 5, 6, 7, 8, 9, 10 Muy importante	Pregunta de estimación (ítem de Likert)	Opinión
P10	Valoración de importancia: Los factores culturales, hábitos alimentarios, la formación y los placeres gastronómicos:	1 Poco importante, 2, 3, 4, 5, 6, 7, 8, 9, 10 Muy importante	Pregunta de estimación (ítem de Likert)	Opinión
P11	En relación con los siguientes dos retos a desarrollar por la ciencia, ¿cuál cree Ud. que es el más importante?	1 Incrementar la producción de alimentos para una mayor disponibilidad en la población, 2 Incrementar la calidad de los alimentos para mejorar su valor nutricional y sabor	Múltiple excluyente	Aspiraciones alimentarias

Fuente: elaboración propia

En referencia al **tipo de preguntas**, en el cuestionario encontramos (Anguita, J. C., 2003):

1. Preguntas de selección múltiple excluyentes: el encuestador, para reflejar su opinión o situación personal, debe elegir una de entre las opciones predefinidas.
2. Preguntas cerradas: el encuestar debe elegir entre dos opciones.
3. Preguntas de estimación: se ofrecen al encuestador como alternativas de respuestas graduadas en intensidad. Concretamente, se utiliza los ítems de Likert² para saber la importancia que se atribuye a los determinados factores/ejes del sistema alimentario.
4. Preguntas semiabiertas: se da la libertad al encuestador para que conteste lo que quiera pero siguiendo uno parámetro de definición de respuesta, como es el caso de la variable Fecha de nacimiento.

Según la **naturaleza del contenido**, en el cuestionario encontramos:

1. Preguntas sociodemográficas, concretamente cinco: Sexo, Fecha de nacimiento, Nivel de estudios, Clase social y Comarca de residencia.
2. Preguntas sobre importancia de un tema, 4 concretamente: una para cada uno de los ejes a estudiar.
3. Preguntas relativas a prácticas y aspiraciones alimentarias, concretamente encontramos dos: tipo de alimentación en la que se encuentra más cómodo y cuál de los retos cree que es mejor que desarrolle la ciencia.

Y finalmente, en referencia a su función, cabe destacar que para tener un buen nivel de rigurosidad y calidad en el cuestionario se preguntó por la fecha de nacimiento en vez de la edad. De esta manera la pregunta tiene una función de control y consistencia, ya que al cruzarla con la comarca y el género, será una mecanismo eficaz para comprobar la presencia o ausencia de duplicados.

3.1.6. Prueba piloto o pre-test

Una vez redactadas las preguntas y adecuadamente situadas en el cuestionario, se realizó una prueba piloto con un grupo de ocho personas. Esto nos permitió determinar si las preguntas

² La escala de Likert, muy utilizada en ciencias sociales y humanidades, es una escala aditiva con un nivel ordinal, constituida por una serie de ítems ante los cuales se solicita la reacción del sujeto (Namakforoosh, 2000)

habían sido correctamente comprendidas por todos los sujetos, si la duración era adecuada y alguna otra deficiencia que hacía referencia a la distribución de importancia entre las posibles opciones de respuesta.

Se introdujeron los cambios y se elaboró el formato definitivo, tanto en papel como online y disponible en catalán y castellano. Finalmente, se procedió a la difusión del cuestionario.

3.2. Técnicas estadísticas utilizadas para la realización de los análisis posteriores

Una vez hecha la difusión y la recogida de datos, se procedió a hacer los análisis. En concreto en este apartado se van a explicar las técnicas cuantitativas que se han utilizado.

3.2.1. Técnica estadística de muestreo: Ponderaciones

Al tratarse de una encuesta de cuotas ponderadas, es muy común que algunas de ellas estén sobredimensionadas o subdimensionadas. Para ello, utilizaremos la técnica estadística de la ponderación que consiste en corregir cualquier desequilibrio en los perfiles de muestra después de la recopilación de datos. De manera que aplica una ponderación a cada uno de los individuos para que, los criterios de ponderación, en nuestro caso Sexo, Edad y Provincia, tengan los pesos del universo estudiado (Cataluña en este estudio).

Para aplicarla en la base de datos y poder llevar a cabo los posteriores análisis, se ha utilizado una librería específica de R: ***“Package ‘weights’”*** que *“proporciona una variedad de funciones para la realización de estadísticas ponderadas simples, como las correlaciones de Pearson ponderados, correlaciones parciales, estadísticas Chi-cuadrado, histogramas, y pruebas t”* (Pasek, J., 2018). De manera que ha permitido hacer tanto el análisis univariante como el bivalente con las ponderaciones aplicadas y así establecer una correcta dimensionalidad de las cuotas.

3.2.2. Técnica estadística sobre la distribución: bootstrapping

Por otro lado, en el análisis bivalente, para comprobar si hay diferencias estadísticamente significativas entre las variables sociodemográficas y los ejes sobre la alimentación se han realizado

t-test de Student, teniendo en cuenta la ponderación de cada individuo y también aplicando la **técnica de bootstrap** ya que las variables no seguían una distribución Normal $\sim (\mu, \sigma)$. Esta técnica esta compuesta de dos etapas para aproximar la distribución muestral desconocida (Badii, M., 2017):

- Primero, utilizar información de la muestra para aproximar la distribución desconocida de los valores de la población y tomar muchas muestras de bootstrap de esta distribución.
- Segundo, se aproxima la distribución muestral desconocida por medio de la distribución de estimaciones de muchas muestras generadas por el bootstrap.

Concretamente, para el análisis bivalente, hemos aplicado los siguientes criterios de los disponibles en la librería *weights* sobre la técnica de bootstrapping (Pasek, J., 2018):

- Bootse: bootse es un parámetro que produce errores estándar de arranque. Se utiliza cuando hay que abordar problemas de heterocedasticidad como es en nuestro caso.
- Bootp: bootp es un parámetro que produce valores p de bootstrapped en lugar de estimar los valores p a partir de los errores estándar. De esta manera nos aseguramos una mayor fiabilidad del parámetro estadístico: p-valor.
- Bootn: bootn es un parámetro que se utiliza para indicar la cantidad de bootstraps que se deben ejecutar para bootse y bootp. En el estudio, el número de computaciones es de 5.000.

Entonces a partir de la aplicación de la ponderación y de la técnica de bootstrap se ha realizado la prueba t-test con la finalidad de observar si hay diferencias significativas entre medias afijadas por las variables sociodemográficas, como hemos comentado anteriormente. Aunque cabe añadir y es de especial relevancia, entender que los análisis hechos de variables cruzadas, se trata de una resultante, es decir, no son representativos por ellos mismos pero al tratarse de encuestados elegidos aleatoriamente y que el global sí que lo es, la derivación de estos análisis representa esta diversidad.

3.2.3. Técnica estadística de imputación de valores perdidos: Algoritmo Múltiple Imputación by Chained Equations (MICE)

Para la imputación de valores perdidos se ha utilizado el algoritmo *Multiple Imputation by Chained Equations* (MICE). Se trata de una técnica estadística que opera bajo el supuesto de que los datos perdidos son *Missing At Random* (MAR) (Azur J., 2011). A partir de este supuesto, “ejecuta una serie de modelos de regresión mediante los cuales cada variable con datos perdidos se modela condicionalmente a las otras variables. Esto significa que cada variable puede modelarse de acuerdo con su distribución, como por ejemplo, variables binarias modeladas usando regresión logística y variables numéricas modeladas usando regresión lineal” (Azur J., 2011).

En nuestro caso, como se trata de una variable numérica, se ha utilizado el método *norm.boot* que realiza una regresión lineal usando *bootstrap* para determinar la distribución e imputar los valores perdidos.

3.2.4. Técnica estadística para el clustering: distancia de Gower, método de ward, Silhouette test y algoritmo Partición Alrededor de Medoids (PAM)

El análisis de clúster es una técnica estadística multivariante que busca agrupar elementos (o variables) tratando de lograr la máxima homogeneidad en cada grupo y la mayor diferencia entre ellos.

Para ello se calcula la matriz de distancia, que en nuestro caso se ha hecho a partir de la distancia de Gower ya que es muy útil para manejar diferentes tipos de variables a partir del coeficiente de disimilitud de Gower (Gower, 1971) y concretamente, la extensión de Kaufman y Rousseeuw.

La diferencia final entre las unidades i y j se obtiene como una suma ponderada de las diferencias para cada variable (Filaire, 2018):

$$d(i, j) = \frac{\sum_k \delta_{ijk} d_{ijk}}{\sum_k \delta_{ijk}} \quad (8)$$

Concretamente para las variables categóricas ordinales, en nuestro caso las cuatro variables sobre la valoración de los ejes, se calcula mediante de la siguiente manera:

$$z_{ik} = \frac{(r_{ik} - 1)}{\max(r_{ik}) - 1} \quad (9)$$

Donde los valores se sustituyen con el índice de posición correspondiente, r_{ik} en los niveles de factores y z_{ik} , se tratan como observaciones de una variable de escala de intervalo.

En el caso de las variables nominales como es la variable sobre el tipo de alimentación o la variable sobre el reto a desarrollar por la ciencia, la manera de calcularla es la siguiente:

$$d_{ijk} = \frac{|x_{ik} - x_{jk}|}{R_k} \quad (10)$$

Siendo $d_{ijk} = 0$ si $x_{ik} = x_{jk}$, 1 y R_k el rango de la k ésima variable.

Una vez calculada la matriz de distancia, se utiliza el método ward para realizar el dendograma que minimizar la varianza total dentro del grupo. Es un método que tiene la tendencia a encontrar grupos compactos y esféricos y se define como el cuadrado de la distancia euclidiana entre puntos:

$$d_{ij} = d(\{X_i\}, \{X_j\}) = \|X_i - X_j\|^2 \quad (11)$$

Después de realizar el dendograma, se utiliza el test Silhouette para validar la consistencia dentro de los grupos de datos, de manera que proporciona una representación gráfica de qué tan bien se ha clasificado cada objeto (Rousseeuw, P., 1987). Se computa de la siguiente manera:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}, \text{ Si } |C_i| > 1 \quad (12)$$

Los valores van de -1 a $+1$, donde un valor alto indica que el objeto está bien emparejado con su propio grupo y mal emparejado con los grupos vecinos.

Y finalmente, para ejecutar el clúster no jerárquico se utiliza el algoritmo de Partición Alrededor de Medoids (PAM) ya que encaja bien con la distancia de Gower (Filaire, T., 2018). Se trata de una técnica de partición de agrupamiento que agrupa el conjunto de datos de n objetos en k grupos conocidos a priori. Concretamente, tiene las siguientes características (Filaire, T., 2018):

- **Pros:** más robusto al ruido y valores atípicos en comparación con k -medias y produce un "individuo típico" para cada grupo (útil para la interpretación).
- **Contras:** consume mucho tiempo y requiere mucha computadora (el tiempo de ejecución y

la memoria son cuadráticos).

4 Depuración, estructuración y presentación de los datos

4.1 Depuración, imputación de valores perdidos, agrupación y codificación de los datos

Como ya comentamos en el apartado 3.1 La encuesta, una de las variables de control es el Año de Nacimiento. Aunque el interés principal era saber la edad, al preguntar por la fecha entera le añadíamos la capacidad de **verificar la no duplicación de encuestados**.

Concretamente, para poder detectarlo se ha cruzado la variable Fecha de Nacimiento, con la variable Sexo y la variable Comarca de residencia. De esta manera, la probabilidad de encontrar dos personas de 419 encuestados que tengan las mismas características y tan concretas como es el día, mes y año exacto de nacimiento, es mínima. Y en efecto, al realizar el análisis se ha comprobado que no hay dos personas con esos mismos rasgos demográficos y que por tanto, podemos afirmar que no hay evidencias para pensar que la misma persona ha contestado dos veces el cuestionario.

Una vez hecha la depuración de posibles duplicados, se ha procedido a imputar, con el algoritmo MICE, los **valores perdidos**. La única variable que presentaba este tipo de valores era la Fecha de Nacimiento debido a que, por defecto, en la aplicación online del cuestionario el año que ponía era el 2019. De manera que hubo 14 encuestados con ese año, los cuales se les aplicó el algoritmo y los valores resultantes son los que se utilizaron para los siguientes análisis.

El tercer paso a realizar fue la **agrupación de dos variables**: la Edad y la Comarca en la cual reside el encuestado. En la primera, la Edad, como se ha podido observar el tramo de 45 o más acoge mucho más años que los otros rangos. Esto se debe a la dificultad de obtener respuestas de gente mayor de 55 años. De todas formas, en los menores de 45 años (al ser edades más fáciles de captar) se ha querido asegurar la diversidad de edades y plasmar mejor la diversidad de la población catalana. Para solucionar esta desproporción se ha agrupado la variable en dos grupos, siguiendo el corte de edad que propone Martín en su artículo: *“adultos jóvenes”* (de 16 a 44) y *“adultos maduros y maduros”* (de 45 y más) (F., Martín, 2005). Y en la segunda, la Comarca de residencia, el tener 41 diferentes sí que ha ayudado a la depuración de duplicados pero para su análisis, era muy complicado al tener tantas categorías. Por eso se ha agrupado en las cuatro

provincias de Cataluña: Barcelona, Tarragona, Lleida y Girona.

Y finalmente, el cuarto paso que se ha realizado es la **codificación de todas las variables categóricas o nominales**: Sexo, Edad, Provincia, Clase social, Estudios, Tipo de alimentación y Visión del papel de la ciencia, que es esencial para aplicar las ponderaciones y las diversas técnicas estadísticas que se harán posteriormente.

4.2. Distribución de las respuestas obtenidas versus las esperadas y aplicación de la ponderación

Después de los meses de difusión del cuestionario y recogida de datos, se ha conseguido un total de 419 respuestas, número superior a las 385 necesarias para tener un tamaño de muestra suficiente para tener resultados significativos a nivel de Cataluña, concretamente representa un 8,83% más de las necesarias. Al desglosarlo por las afijaciones presentadas en el apartado anterior, vemos que, en valores absolutos, se distribuyen de la siguiente manera [Tabla 7]:

Tabla 7. DISTRIBUCIÓN DE LAS RESPUESTAS OBTENIDAS Y ESPERADAS POR AFIJACIONES Y PORCENTAJE QUE REPRESENTA

Hombres	15-24	25-34	35-44	De 45 o más	TOTAL
Cataluña	26; 22 (118,2%)	28; 28 (100,0%)	38; 38 (100,0%)	79; 99 (79,8%)	171; 187 (91,4%)
Barcelona	19; 17 (111,8%)	20; 20 (100,0%)	34; 28 (121,4%)	59; 72 (81,9%)	132; 137 (96,4%)
Girona	1; 2 (50,0%)	4; 3 (133,3%)	1; 4 (25,0%)	11; 10 (110,0%)	17; 19 (89,5%)
Lleida	0; 1 (0,0%)	3; 2 (150,0%)	2; 2 (100,0%)	3; 6 (50,0%)	8; 11 (72,7%)
Tarragona	6; 2 (300,0%)	1; 3 (33,3%)	1; 4 (25,0%)	6; 11 (54,5%)	14; 20 (70,0%)

Mujeres	15-24	25-34	35-44	De 45 o más	TOTAL
Cataluña	41; 22 (186,4%)	54; 27 (200,0%)	50; 36 (138,9%)	103; 113 (91,2%)	248; 198 (125,3%)
Barcelona	36; 16 (225,0%)	36; 20 (180,0%)	44; 27 (163,0%)	85; 84 (101,2%)	201; 147 (136,7%)
Girona	3; 2 (150,0%)	8; 3 (266,7%)	4; 4 (100,0%)	11; 11 (100,0%)	26; 20 (130,0%)
Lleida	0; 1 (0,0%)	5; 1 (500,0%)	0; 2 (0,0%)	1; 6 (16,7%)	6; 10 (60,0%)
Tarragona	2; 2 (100,0%)	5; 3 (166,7%)	2; 4 (50,0%)	6; 12 (50,0%)	15; 21 (71,4%)

*Contenido de la celda: (Nº de respuestas observadas); (Nº de respuestas esperadas) ;
(Porcentaje que presenta: N° resp observ./N° resp esper)

Fuente: elaboración propia

En la tabla observamos el número de respuestas observadas (primer número de la celda) frente al número de respuestas esperadas (segundo número de la celda) y el porcentaje que presenta entre paréntesis. En primer lugar, si nos fijamos en la variable **sexo**, vemos como en las mujeres tenemos el total de respuestas esperadas (231 de 198 que necesitábamos), mientras que en los hombres se ha alcanzado el 91,4% (171 de 187). En segundo lugar, al mirarlo por la variable **provincia** encontramos que en Barcelona se ha alcanzado la cuota (333 observadas frente a 284 esperadas). De la misma manera, la provincia de Girona con un 110,3% (43 respuestas de las 39 que se requerían). En cambio, en el caso de Tarragona se ha logrado un 70,7% de las respuestas esperadas (29 de 41) y finalmente, Lleida, la provincia menor representada aunque con un porcentaje mayor del 50%, concretamente del 66,67% (14 respuestas de las 21 que se esperaban) [Tabla 7].

Y si lo analizamos por la variable **edad** se observan cuatro franjas diferentes. La primera se compone de los que tienen de entre 15 y 24 años. Se trata de una categoría sobredimensionada ya que se ha logrado un 152,3% de respuestas de las que se necesitaban (67 de 44). El segundo rango va de 25 a 34 años y se ha obtenido un 149,1% (82 de 55), por tanto, se ha conseguido la cuota esperada. El tercero es el que va de 35 a 44 años y se ha alcanzado un 118,9% (88 de 74), con lo cual se trataría de un rango igualmente bien representado. Finalmente, el último rango que encontramos es el que va de 45 años o más. En éste hemos obtenido un 85,9% (182 de 212). Por otro lado, al realizar los análisis se tendrá en cuenta la muestra que tenemos en cada grupo de edad ya que son muy diferentes entre ellos (punto argumentado en el apartado anterior, mientras que de 15 a 24 tenemos un número de respuestas esperadas de 44 y de 25 a 34 años un número de 55 respuestas, en el caso de 35 a 44 años de edad es de 112 y en el rango de más de 45 años es de 175. Por tanto, para aplicar diferentes técnicas estadísticas, se aplicará la agrupación de la edad explicada anteriormente.

A partir de este primer análisis, se puede observar como hay algunas cuotas que están sobredimensionadas: mujer en el caso del sexo, Barcelona y Girona en el caso de la provincia y los rangos de 15-24 años, 25-34 años y 35-44 años en la variable edad. De manera que al ponderar, no se pierde la riqueza de los datos de aquéllos que sobrepasan al número de encuestados esperados sino que simplemente, le damos un valor menor, concretamente el que le corresponde en comparación a la población real. Por otro lado, también encontramos cuotas subdimensionadas: hombre en la variable sexo, Lleida y Tarragona en la provincia y el rango de más de 45 años en el caso de la edad.

Teniendo en cuenta estos tres criterios demográficos, se han imputado las ponderaciones correspondientes [Tabla 8]. Como se observa aquí la edad ya ha sido agrupada en dos grupos: “Jóvenes y Adultos jóvenes” y “Adultos mayores y Mayores”. Por lo que los análisis posteriores ya se presentarán con esta agrupación, sin dejar de tener en cuenta el poder representativo en tiene el grupo “Jóvenes y Adultos jóvenes” ya que procede de tres rangos que han sido cuota a tener en cuenta en las afijaciones.

Tabla 8. PONDERACIONES POR AFIJACIONES DE LAS VARIABLES SEXO, EDAD Y PROVINCIA

Hombres	Jóvenes y Adultos jóvenes (16 a 44 años)	Adultos mayores y Mayores (De 45 o más años)
Cataluña	0,957	1,253
Barcelona	0,890	1,220
Girona	1,500	0,909
Lleida	1,000	2,000
Tarragona	1,125	1,833

Mujeres	Jóvenes y Adultos jóvenes (16 a 44 años)	Adultos mayores y Mayores (De 45 o más años)
Cataluña	0,586	1,097
Barcelona	0,543	0,988
Girona	0,600	1,000
Lleida	0,800	6,000
Tarragona	1,000	2,000

Fuente: elaboración propia

Por tanto, lo que vemos en esta tabla es el peso que tiene cada persona encuestada dependiendo de sus características sociodemográficas. Por ejemplo una mujer de Barcelona de 16 a 44 años tendrá un peso de la mitad, mientras que por ejemplo el hombre de 45 años o más de Lleida valdrá el doble en la encuesta. Por lo que, utilizando esta técnica estadística, se tiene que tener cuidado a no tener coeficientes muy altos, ya que significaría que una individuo está representando una gran parte de la población y podríamos caer en conclusiones erróneas. En nuestro estudio vemos como, en el caso de los hombre, lo máximo que puede valer es el doble, por lo que no presentaría problemas muy relevantes. Y en el caso de las mujeres, el número llega hasta seis, es decir, una mujer de Lleida vale por seis, lo que nos avisa que las conclusiones a las que se puedan llegar con este grupo pueden ser equivocadas.

En todo caso, esta técnica nos ayuda a conocer y analizar la realidad de Cataluña de una manera mucho más precisa.

5 Resultados

5.1 Estadísticos descriptivos univariantes

Después de aplicar las ponderaciones mostradas en el apartado anterior para cada uno de los individuos de la encuesta, se procede a mostrar cuál es el perfil de la población catalana sobre la importancia de los diferentes aspectos del sistema alimentario estudiados. Para ello, se han calculado estadísticos descriptivos univariantes que permitirán caracterizar y tener una visión general de los datos recogidos en el cuestionario.

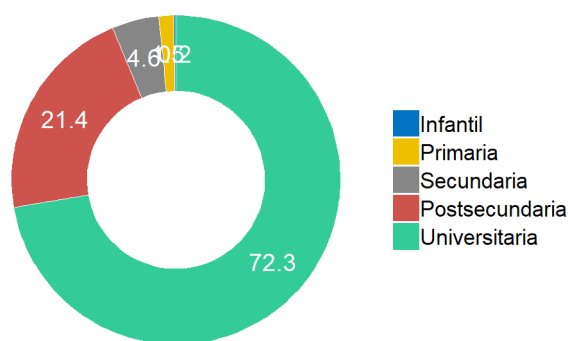
5.1.1. Variables sociodemográficas

En los siguientes gráficos se puede observar cómo se distribuyen las dos variables sociodemográficas: Nivel de estudios y clase social percibida, ya que son las que no han sido criterio de ponderación. En el caso de las otras, el resultado sería el mismo del como se distribuye el universo de la población en Cataluña.

Por una lado, en el caso de los **estudios** [figura 3], lo que podemos destacar es el gran porcentaje de encuestados que tienen estudios universitarios, concretamente un 75,8% (278 encuestados). Esto es un punto a tener en cuenta a la hora llevar a cabo posteriores análisis y/o conclusiones sobre esta variable. De hecho, en Cataluña solamente un 32,4% tiene estudios superiores (INE, 2019), con lo cual nuestra muestra está lejos, en este aspecto, de representar el universo. Aunque, al no ser considerada una variable de ponderación, la implicación que puede tener en el estudio no es de gran relevancia.

Por otro lado, encontramos un 21,4% de encuestados (82) que han alcanzado estudios Post-secundarios, mientras que un 4,6% (18 encuestados) tienen hasta secundaria, un 1,5% (6 encuestados) primaria y un 0,2% (1 encuestado) tiene infantil.

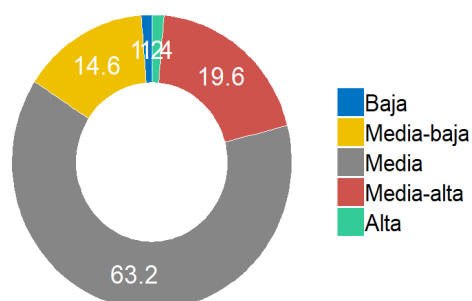
Figura 3. DISTRIBUCIÓN DE RESPUESTAS A LA PREGUNTA 4. ¿CUÁL ES SU NIVEL DE ESTUDIOS?



Fuente: elaboración propia

Y por otro lado, la pregunta sobre la clase social: “Si tuviera que elegir una clase social, a cuál diría que pertenece?” tenemos que tener en cuenta que se trata de una percepción que tiene el individuo, no de un indicador objetivo medido y que pueda ser comparativo. Como vemos en el gráfico [figura 4], vemos que el 63,2% (243 encuestados) son de clase media. Seguidamente, encontramos un 19,6% (76 encuestados) de clase Media-Alta y un 14,6% (56 encuestados) de clase Media-Baja Y, por otro lado, 2,7% y un 2,4% consideran que pertenecen a una clase social Alta y Baja, respectivamente. Esto si lo comparamos con los datos oficiales del INE, se observa como hay una fuerte similitud, ya que exponen que el 59,2% de la sociedad española pertenece a la clase media, el 21,4 en la clase Alta /Media-alta y que el 19,3% se encuentra en la clase Media-baja/Baja (INE, 2019). Por tanto, hay evidencias para afirmar que hay una buena representatividad de la muestra en este aspecto.

Figura 4. DISTRIBUCIÓN DE RESPUESTAS A LA PREGUNTA 5. SI TUVIERA QUE ELEGIR UNA CLASE SOCIAL, ¿ A CUÁL DIRÍA QUE PERTENECE?



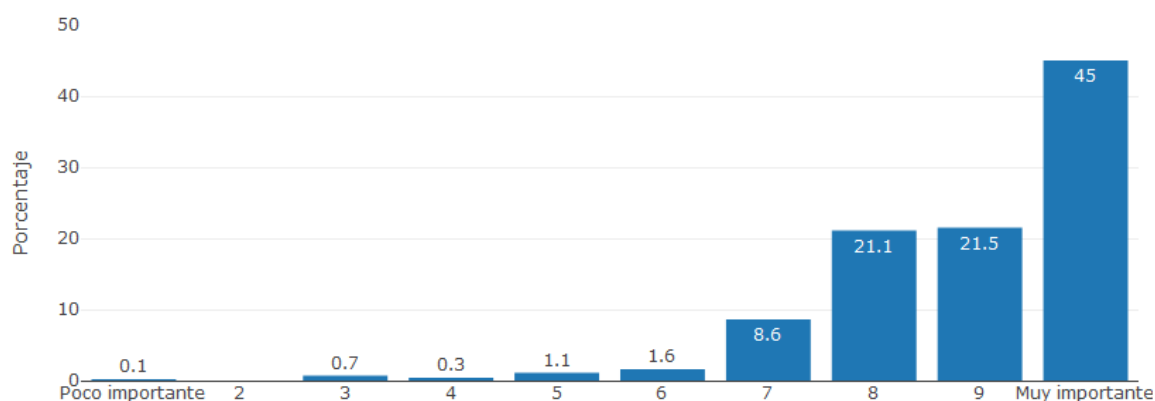
Fuente: elaboración propia

5.1.2. Variables de valoración de la importancia de los cuatro ejes: disponibilidad alimentaria, economía, políticas públicas y conocimiento/saber del sistema alimentario

Una vez aplicados los criterios de ponderación según los tres criterios demográficos (edad, sexo y provincia), se procede a ver como puntúa la sociedad catalana el sistema alimentario. Los siguientes gráficos muestran la distribución que siguen las cuatro preguntas que hacen referencia a los ejes definidos sobre el tema: disponibilidad de alimentos, economía de la sociedad para adquirir los alimentos, las políticas públicas que hacen referencia a la alimentación y la seguridad alimentaria y finalmente, el cuarto, el conocimiento y los aspectos culturales que tiene la sociedad sobre el tema.

- **Disponibilidad:** los datos reflejan como los encuestados creen que la disponibilidad alimentaria es un aspecto importante. Concretamente, encontramos una media de 8,9 puntos de importancia. Observamos como los porcentajes son casi nulos hasta llegar a una puntuación de 7 que encontramos un 8,6% (33 encuestados). Seguidamente, con un porcentaje muy igual, del 21,1% (81 encuestados) y del 21,5% (83 encuestados) encontramos los que lo valoran con un 8 y 9, respectivamente. Y finalmente, el gran pico lo encontramos en los que le dan mucha importancia a la disponibilidad de alimentos, con un 45% (173 encuestados).

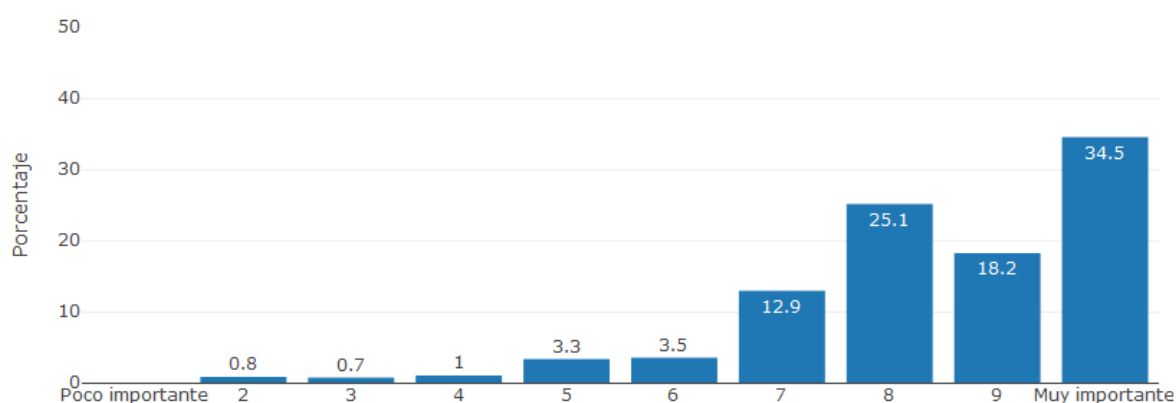
Figura 5. DISTRIBUCIÓN DE RESPUESTAS A LA PREGUNTA DE VALORACIÓN A: 7. TENER LA DISPONIBILIDAD DE ACCESO A ALIMENTOS NUTRICIONALMENTE ADECUADOS A LA DIETA



Fuente: elaboración propia

- **Economía.** En este caso, la media es levemente más baja: 8,45 puntos. Podemos ver como sí que se considera importante, aunque a diferencia de la disponibilidad alimentaria, destacan dos aspectos: uno, que las frecuencias están distribuidas de manera más gradual a partir de una importancia de 6 puntos. Concretamente, con una puntuación de 7 hay un 12,9%, con un 8 un 25,1%, con un 9 un 18,2% y un 34,5% (133 encuestados), con el porcentaje más alto, considera el nivel económico es muy importante (10 puntos). Y dos, que el aumento que hay a partir de la valoración de 6 no es continuo ya que encontramos un decrecimiento en los que lo valoran con una importancia de 9 puntos.

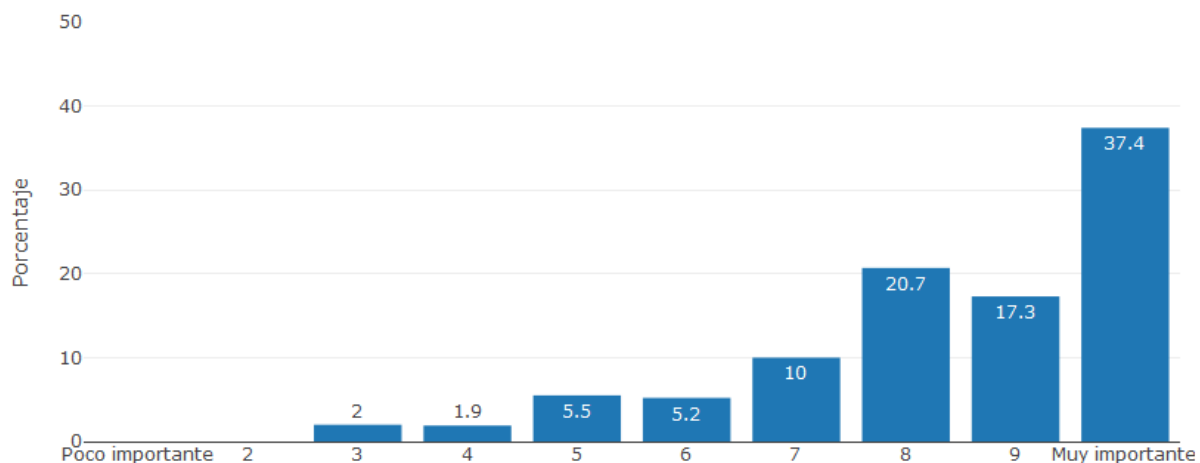
Figura 6. DISTRIBUCIÓN DE RESPUESTAS A LA PREGUNTA DE VALORACIÓN A: 8. TENER EL NIVEL ECONÓMICO PARA COMPRAR LOS ALIMENTOS NECESARIOS



Fuente: elaboración propia

- **Políticas públicas:** se observa una media de valoración de importancia del 8,37, un poco menos que el eje de economía. Concretamente, vemos como el 37,4% (144 encuestados) piensa que es muy importante (10 puntos), seguido de un 20,7% (80 encuestados) que le dan una importancia de 8 y de un 17,3% por quienes lo valoran en 9 puntos. Con porcentajes más bajos encontramos un 10% por lo que le dan 7 puntos de importancia, y del 5,5 y 5,2 los que valoran las políticas con 5 y 6 puntos respectivamente. Por tanto, se trata, en comparación al eje de disponibilidad alimentaria y el de economía, el que tiene menor media y menores porcentajes en las valoraciones más altas (8, 9 y 10).

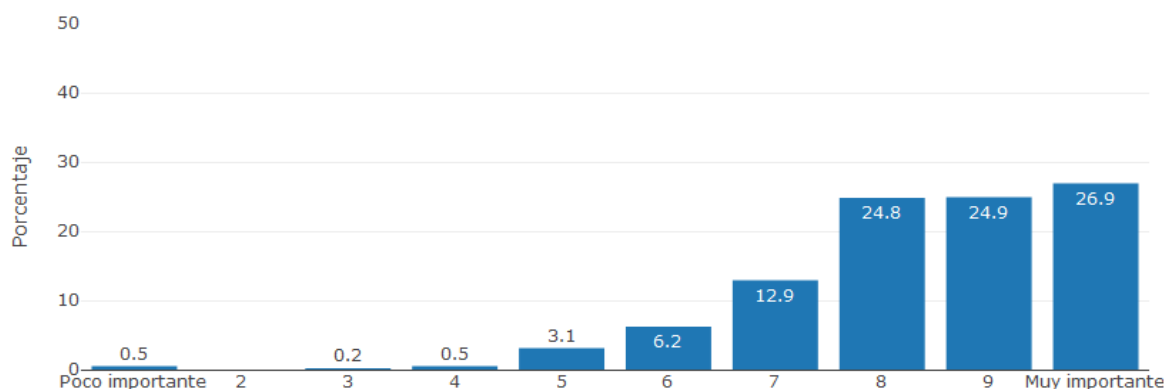
Figura 7. DISTRIBUCIÓN DE RESPUESTAS A LA PREGUNTA DE VALORACIÓN A: 9. LAS POLÍTICAS PÚBLICAS PARA FAVORECER LA PRODUCCIÓN DE ALIMENTOS Y LA SEGURIDAD ALIMENTARIA



Fuente: elaboración propia

- **Factores culturales.** Finalmente aquí se observa como la importancia de los factores culturales se valora con los encuestados con una media, igual que el eje de políticas públicas, de 8,37 puntos. La gran diferencia es que encontramos, al contrario que el anterior eje, una homogeneidad entre las puntuaciones 8, 9, 10, con una porcentajes de 24,8% (95 encuestados), 24,9% (96 encuestados) y 26,9% (104 encuestados) respectivamente. A partir del una valoración de 7 puntos, los porcentajes decrecen a medida que las puntuaciones son más bajas.

Figura 8. DISTRIBUCIÓN DE RESPUESTAS A LA PREGUNTA DE VALORACIÓN A: 10. LOS FACTORES CULTURALES, HÁBITOS ALIMENTARIOS, LA FORMACIÓN Y LOS PLACERES GASTRONÓMICOS



En conclusión, podemos decir que en general vemos como las valoraciones en los cuatro ejes son relativamente altas, teniendo el 75% de las respuestas en 8 o más puntos de importancia. Y por otro lado, si vemos la jerarquía entre las medias de los cuatro ejes vemos que queda de la

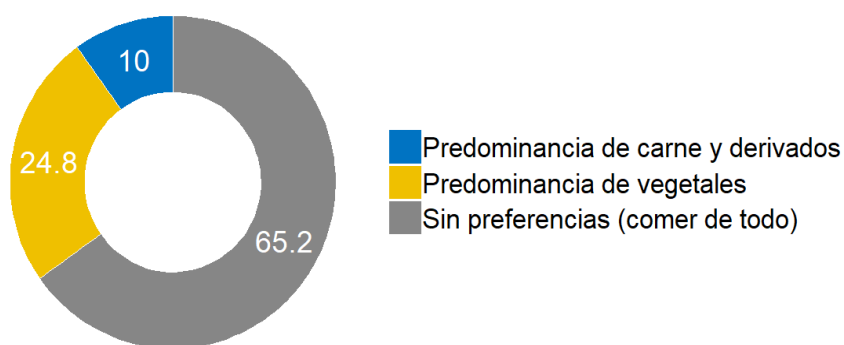
siguiente manera: en primer lugar la disponibilidad de alimentos con un media del 8,9, en segundo lugar las políticas públicas que tiene el estado sobre el tema de la alimentación con una media de 8,5, en el tercero hay la economía con un 8,47 de media y finalmente, en el último lugar, los factores culturales que tiene la sociedad al respecto con una media del 8,37.

5.1.3 Variables sobre práctica y aspiración en el sistema alimentario

En el siguiente gráfico [figura 9] se puede observar cómo se distribuyen los encuestados en función del tipo de alimentación que se encuentra más cómodo. Esta pregunta tiene como finalidad, saber de una manera más exacta la variable S_2 del eje “Saber”, que hace referencia al “comportamiento de la dieta” y que se midió en las investigaciones anteriores a partir de los kilogramos de consumo de carne.

En este caso, encontramos que el 65,2% (251 encuestados) de la población catalana no tiene preferencias en la dieta, mientras que un 24,8% (95 encuestados) se siente más cómodo en un tipo de alimentación donde hay predominancia de vegetales y un 10% (39 encuestados) prefiere la dieta basada en carne y derivados.

Figura 9. DISTRIBUCIÓN DE RESPUESTAS A LA PREGUNTA 6. EN QUE TIPO DE ALIMENTACIÓN SE ENCUENTRA MÁS COMODO

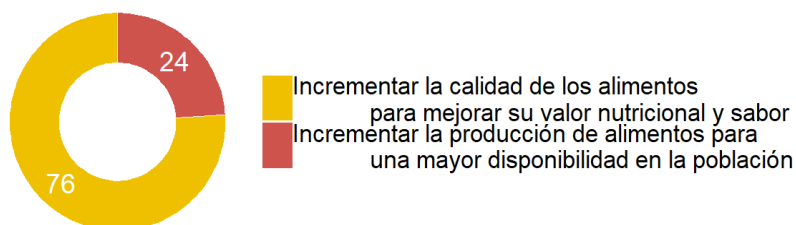


Fuente: elaboración propia

Y finalmente, en el siguiente gráfico [figura 10] se pueden observar los resultados de cómo prefieren enfocar los retos de la ciencia los catalanes en el tema de la alimentación. La gran mayoría, concretamente un 76% (293 encuestados) prefiere incrementar la calidad de los alimentos para mejorar su valor nutricional y sabor, mientras que el 24% (92 encuestados) opina que es mejor incrementar la producción de alimentos para una mayor disponibilidad alimentaria

en la población.

Figura 10. DISTRIBUCIÓN DE RESPUESTAS A LA PREGUNTA 11. EN RELACIÓN CON LOS SIGUIENTES RETOS A DESARROLLAR POR LA CIENCIA, ¿CUÁL CREE UD. QUE ES EL MÁS IMPORTANTE?



Fuente: elaboración propia

5.2 Análisis estadístico bivalente entre los 4 ejes y las variables sociodemográficas: Sexo, Edad y Provincia

Este apartado tiene como objetivo realizar un análisis bivalente entre las variables sociodemográficas que hayan sido criterio de ponderación (sexo, edad y provincia) con las cuatro preguntas relativas a los ejes: disponibilidad alimentaria, economía, política y factores culturales para poder identificar cuál es el perfil sociodemográfico que determina los ejes del sistema alimentario.

- Variable “**tener *disponibilidad* de acceso a alimentos**”

Como vemos en la siguiente tabla [Tabla 9], vemos que las variables que presentan medias estadísticamente significativas ($p\text{-value} < 0,05$), es decir, que aceptamos la hipótesis alternativa (H1) de que los grupos son diferentes entre categorías es el sexo. Concretamente, observamos cómo las mujeres le dan una mayor importancia a la disponibilidad de alimentos nutricionalmente adecuados a la dieta, con una media de 9,15, mientras que los hombres le dan una importancia de media de 8,65. Igual de importante es ver cómo no observamos diferencias significativas entre edad y provincia, es decir que, ni el ciclo de vida en el que se encuentra la persona ni en el territorio donde vive determina la valoración que se da en este eje.

Tabla 9. T-TEST PONDERADO NO PARAMÉTRICO DE LA VARIABLE DISPONIBILIDAD ALIMENTARIA CON LAS SOCIODEMOGRÁFICAS: SEXO, EDAD Y PROVINCIA

Tener la disponibilidad de acceso a alimentos nutricionalmente adecuados a la dieta			
Variable	Categoría	Media	Significación
Sexo	Hombre	8,65	0,0002**
	Mujer	9,15	
Edad	---- no sig----	----	0,1719
Provincia	---- no sig----	----	0,7516

Fuente:elaboración propia

Nota: umbrales de significación: **5%; *10%

- Variable “tener el nivel **económico** para comprar los alimentos necesarios”

La siguiente tabla muestra como al fijarnos en la importancia de tener el nivel económico para comprar los alimentos necesarios para nuestra dieta, no hay diferencias entre perfiles, lo que se traduce que toda la población catalana le da la misma importancia a tener dinero para adquirir alimentos.

Tabla 10. T-TEST PONDERADO NO PARAMÉTRICO DE LA VARIABLE ECONOMÍA CON LAS SOCIODEMOGRÁFICAS: SEXO, EDAD Y PROVINCIA

Variable	Categoría	Media	Significación
Sexo	---- no sig----	----	0,1364
Edad	---- no sig----	----	0,5582
Provincia	---- no sig----	----	0,1083

Fuente:elaboración propia

Nota: umbrales de significación: **5%; *10%

- Variable sobre la importancia de “las **políticas públicas** para favorecer la producción de alimentos y la seguridad alimentaria”

En el caso de esta variable, encontramos un perfil sociodemográfico más determinista ya que hay diversas categorías que han salido estadísticamente significativas. Se observa como el sexo, hombre/mujer, se diferencia entre ellos ($p\text{-value} < 0,05$): las mujeres le dan más importancia a las políticas públicas que los hombres con una media de casi 0,7 puntos más (Media en las mujeres de 8,7 /Media de los hombres de 8,04). Por otra parte vemos como en dos provincias aceptamos la H1, es decir, que son diferentes con el resto: Barcelona y Girona. En el caso de la primera, Barcelona, encontramos que dan una menor importancia a las acciones de políticas públicas en alimentación (con una media de 8,02) que el resto de los catalanes (con una media de 7,92) y, en cambio, Girona, le da una mayor importancia, concretamente una media del 8,43 comparado con

el resto que es de 7,92. Aunque hay que tener en cuenta que la muestra en el caso de Girona es relativamente baja (39 encuestados) y que por tanto, las conclusiones no tienen el carácter de determinantes.

Tabla 11. T-TEST PONDERADO NO PARAMÉTRICO DE LA VARIABLE POLÍTICAS PÚBLICAS CON LAS SOCIODEMOGRÁFICAS: SEXO, EDAD Y PROVINCIA

Las políticas públicas para favorecer la producción de alimentos y la seguridad alimentaria:			
Variable	Categoría	Media	Significación
Sexo	Hombre	8,04	0,0001**
	Mujer	8,7	
Edad	---- no sig----	----	0,5326
Provincia	Barcelona	8,02	0,0283**
	Resto	8,5	
	Girona	8,43	0,0697*
	Resto	7,92	
	Lleida	-- no sig--	0,4909
	Tarragona	-- no sig--	0,1166

Fuente:elaboración propia

Nota: umbrales de significación: **5%; *10%

- Variable sobre la importancia de “los **factores culturales**, hábitos alimentarios, formación y placeres gastronómicos”

Y finalmente, en el caso de la valoración de los factores culturales, ya sea hábitos alimentarios como formación o placeres gastronómicos, vemos cómo la edad es estadísticamente significativa ($p\text{-value} < 0,05$): los catalanes de 45 o más años le dan una importancia mayor que los que tienen entre 16 y 44 años (Media de los “jóvenes y adultos jóvenes” igual a 8,21 /Media de los “adultos mayores y mayores” iguala 8,5).

Tabla 12. T-TEST PONDERADO NO PARAMÉTRICO DE LA VARIABLE FACTORES CULTURALES CON LAS SOCIODEMOGRÁFICAS: SEXO, EDAD Y PROVINCIA

Los factores culturales , hábitos alimentarios, la formación y los placeres gastronómicos:			
Variable	Categoría	Media	Significación
Sexo	---- no sig----	----	0,1625
Edad	Jóvenes y Adultos jóvenes (16 a 44 años)	8,21	0,0461**
	Adultos mayores y Mayores (45 o más años)	8,5	
Provincia	---- no sig----	----	0,5734

Fuente:elaboración propia

Nota: umbrales de significación: **5%; *10%

5.3 Análisis de clúster

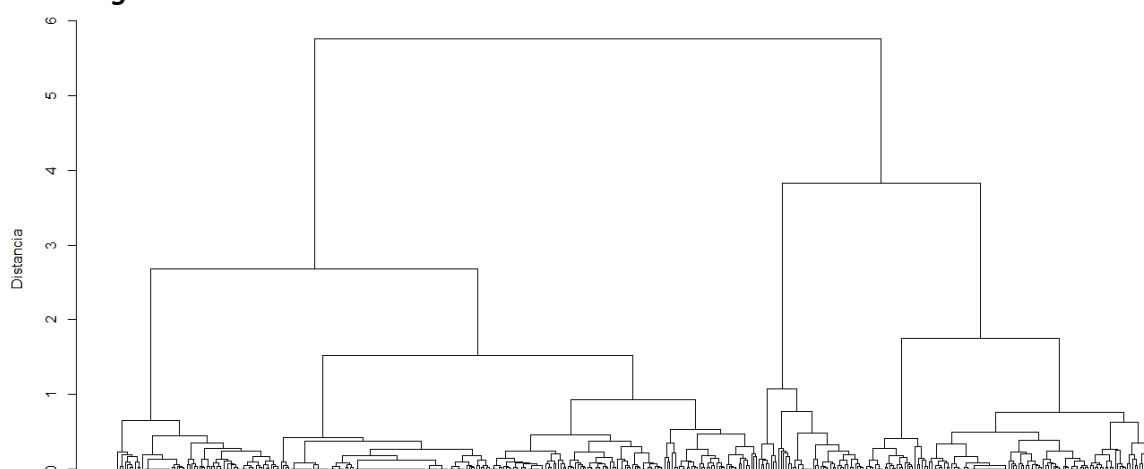
En este apartado se pretende reducir las dimensiones de nuestras variables mediante la técnica de clúster y poder ver más claramente cuáles son las diferencias entre los catalanes respecto al sistema alimentario. Como se ha explicado más detalladamente en el apartado de la metodología, se computa la matriz de distancias a través del calculo de la distancia de Gower y el método de Ward para computar el dendograma.

5.3.1 Clúster jerárquico

En el siguiente gráfico [figura 11] se muestra el dendograma a partir de las variables a estudiar: los 4 ejes (disponibilidad alimentaria, economía, políticas públicas y factores culturales), el tipo de alimentación recodificada como dummy: “sin preferencias: come de todo”, “predominancia de vegetales” y “predominancia de carne” y la variable que nos da información sobre qué papel creen que debería desarrollar la ciencia: “Incrementar la producción de alimentos para una mayor disponibilidad en la población” o “Incrementar la calidad de los alimentos para mejorar su valor nutricional y sabor”.

Vemos, a partir de este primer análisis de cluster jerárquico, que podemos distinguir entre dos, tres, cuatro o seis grupos, dependiendo de cuánto determinístico o global queremos que sea nuestro análisis.

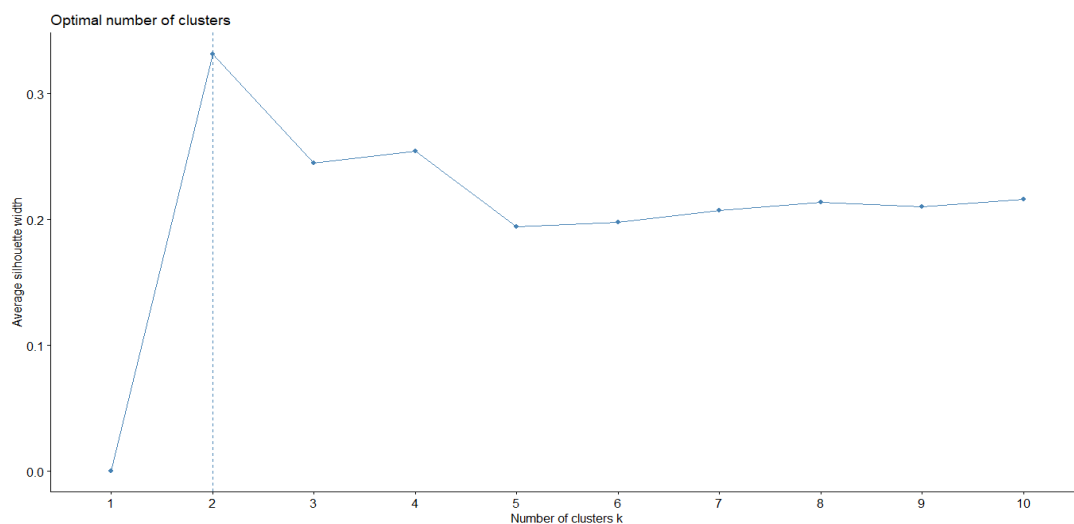
Figura 11. DENDOGRAMA CON MÉTODO WARD Y LAS VARIABLES PONDERADAS



Fuente: elaboración propia

Para determinarlo de una forma rigurosa y consistente, aplicamos el estadístico Silhouette. Al observar cuán similar es un objeto a su propio grupo (cohesión) en comparación con otros grupos (separación), vemos que cuando dividimos los encuestados en dos grupos es cuando tenemos el valor más alto: 0,33, lo cual significa que es donde tenemos un mejor emparejamiento con su propio grupo. Luego, el valor se reduce cuando lo dividimos en 3 (Silhouette = 0,24) y vuelve incrementarse al tener 4 grupos (Silhouette = 0,25). Y a partir de 5 grupos, el valor se estabiliza. De manera que la determinación de grupos se reduce en dos o cuatro. Al ser dos un número de grupos muy global, en el cuál perderemos mucha información, se decide que cuatro es el número ideal de clústers.

Figura 12. TEST SILHOUETTE PARA ANALIZAR LA COHESIÓN ENTRE GRUPOS



Fuente: elaboración propia

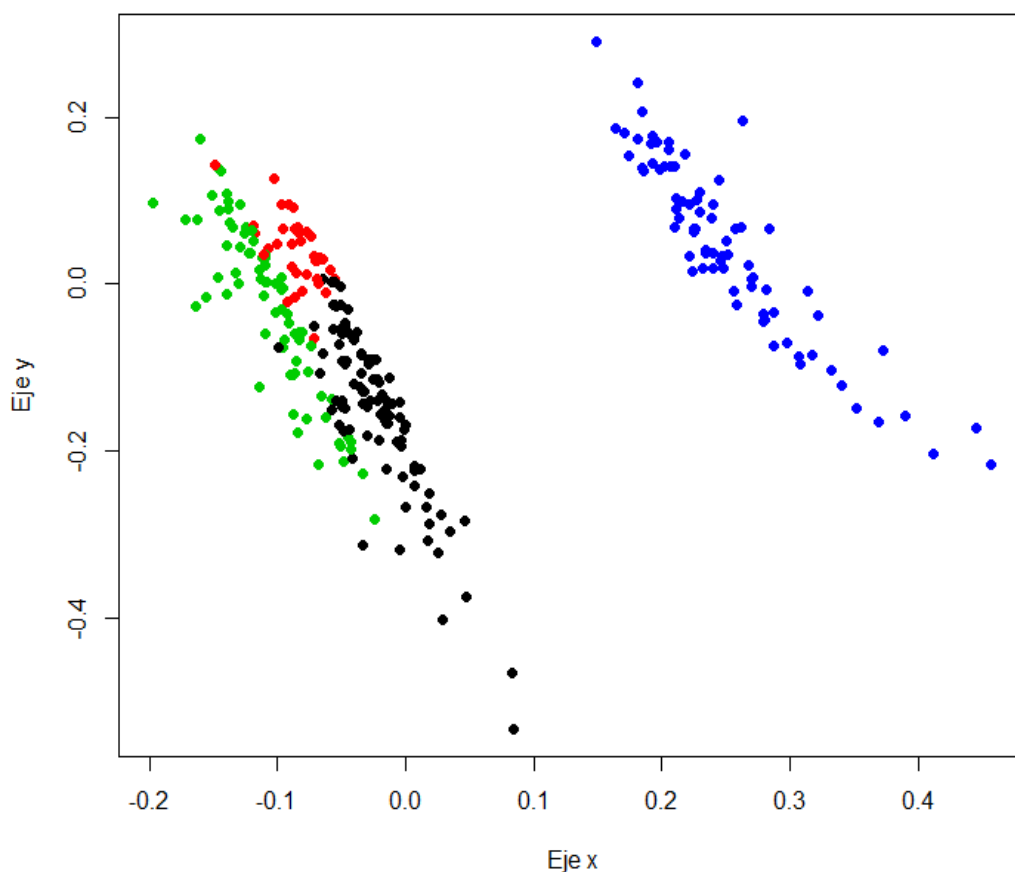
5.3.2 Clúster no jerárquico

Una vez identificados el número de clústers, vamos a observar en cuál se sitúa cada uno de los individuos. Para ello aplicamos una ponderación a cada variable, concretamente en los cuatro ejes se le otorga un valor igual a la media aritmética obtenida en el análisis descriptivo univariante: del eje disponibilidad alimentaria es igual a 8,9, de economía 8,45, de políticas públicas 8,37 y de factores culturales 8,37. Se escala a [0,1; 1], para que tenga cómo máximo una ponderación de 1. En el caso de la pregunta que hace referencia a los retos a desarrollar por la ciencia se le aplica una ponderación igual a 1 y en la dummy sobre tipo de alimentación, al tener tres variables (una para cada categoría) se le aplica un 0,33 para cada una, de manera que la suma sea igual a 1. El vector

resultante es: $c = (0.89, 0.845, 0.837, 0.837, 1, 0.333, 0.333, 0.333)$.

Teniendo en cuenta como punto inicial el vector de la solución de la agrupación jerárquica en $k = 4$ grupos y las ponderaciones de las variables sociodemográficas (sexo, edad y provincia) para cada individuo, se ha llevado a cabo la computación del clúster. Y finalmente para poderlo representar se ha realizado un escalamiento multidimensional de dos dimensiones y el resultado es el siguiente [figura 13]:

Figura 13. VISUALIZACIÓN DE CÓMO SE DISTRIBUYEN LOS CUATRO CLÚSTERS



Fuente: elaboración propia

A partir de las correlaciones entre las variables y los vectores del escalamiento multidimensional, podemos interpretar los ejes y consecuentemente identificar las características de cada grupo. Para ello identificamos correlaciones más altas en valor absoluto. Por un lado, en el eje x podemos ver que es la variable que hace referencia a los retos a desarrollar por la ciencia con una correlación positiva de 0,934. Por tanto, en el lado derecho del eje x se encuentran los individuos que tienen una tendencia a pensar que se tiene que “Incrementar la calidad de los alimentos para

mejorar su valor nutricional y sabor”, mientras que a la izquierda hay lo que tienden a pensar que es mejor *“Incrementar la producción de alimentos para una mayor disponibilidad en la población”*. Por otro lado, en el eje y encontramos que las correlaciones, en valores absolutos, más altas son las que hacen referencia en el eje de la disponibilidad alimentaria, economía y políticas públicas en el ámbito del sistema alimentario. De manera que, en la parte superior hay aquellos que tienen una tendencia en valorar cómo importante estos ejes mientras que los de abajo, lo valoran como menos importante.

Tabla 13. CORRELACIONES ENTRE LOS EJES Y LAS VARIABLES DEL CLÚSTER

Variables	Eje x	Eje y
Disponibilidad	-0,33	0,693
Economía	-0,041	0,643
Políticas públicas	-0,345	0,709
Factores culturales	-0,223	0,563
Reto a desarrollar por la ciencia	0,934	0,312
Sin preferencias: como de todo	0,175	-0,007
Predominancia de vegetales	-0,182	0,146
Predominancia de carne	-0,012	-0,206

Fuente: elaboración propia

A partir de este análisis, vamos a caracterizar los 4 grupos. El azul se trata de individuos que por un lado, le dan importancia a los ejes de disponibilidad alimentaria, economía y a las políticas públicas y a la vez dan prioridad al incremento en la calidad de los alimentos en lugar del incremento de la cantidad de su producción. Por tanto, estaríamos hablando de un grupo que lo podríamos definir como **“Bienestante”**, con ideas basadas en la sostenibilidad pero teniendo en cuenta que parten de una idea enfocada en su sociedad, su cultura, caracterizada por tener suficiencia alimentaria, tanto en términos de economía, disponibilidad alimentaria o estabilidad política a través del estado de bienestar. Y que este bienestar se tiene que mantener a través de una seguridad económica en la adquisición de alimentos, en la disponibilidad y en la creación de políticas públicas que aseguren el sistema alimentario. En cambio, el grupo verde y el grupo rojo se caracterizan por priorizar el incremento de producción de alimentos para una mayor disponibilidad en la sociedad, aunque mientras que el rojo destaca por dar mucha importancia a los tres ejes que hemos comentado anteriormente, el verde no viene definido por ello, es decir, encontramos tanto valoraciones altas como bajas en la escala de importancia. De manera que el grupo rojo lo definiríamos como **“Consciente/Motivado”** ya que se trata de individuos que le dan importancia a

la pobreza mundial, tienen consciencia de que no todo el mundo tiene la cantidad suficiente de alimentos y lo priorizan antes que incrementar el sabor o el valor nutricional. El grupo verde lo definimos como “**Frustrado/Enfadado**”, aunque le da importancia a la alimentación suficiente de todo el planeta, son los que les dan menos importancia a los mecanismos disponibles para solucionarlo: economía, disponibilidad alimentaria y/o políticas públicas. Y finalmente, el grupo negro lo configuran aquellos que les dan una menor importancia a los tres ejes y que no se definen en cuál tiene que ser el reto a desarrollar por la ciencia, por tanto se trataría de un grupo que lo podríamos llamar como “**Indefinido/Apático**”.

Ahora bien si nos fijamos en los perfiles sociodemográficos de cada clúster vemos un resultado interesante: por un lado, ser mujer o hombre es estadísticamente significativo con pertenecer a un clúster u otro y por otro lado, la provincia también. Es decir, vivir en Barcelona, en Girona, en Lleida o en Tarragona determina pertenecer en un determinado clúster [tabla 14]. En el caso de la edad, no hemos encontrado diferencias estadísticamente significativas.

Tabla 14. DIFERENCIAS ESTADÍSTICAMENTE SIGNIFICATIVAS ENTRE LAS VARIABLES SOCIODEMOGRÁFICAS Y LOS CLÚSTERS

Variables	P-value
Sexo	0,00035**
Edad	0,56436
Provincia	0,00002**

Fuente: elaboración propia

Entonces pasamos a ver cuáles son las diferencias que hay dentro de los grupos. Concretamente en el caso del sexo, vemos que hay diferencias significativas en el grupo Frustrado/Enfadado (grupo verde) y del grupo Occidental (grupo azul): mientras que en el primero predominan las mujeres, en el segundo son los hombres. Por otro lado, si nos fijamos en la provincia, se observa que vivir en Barcelona presenta diferencias significativas. Concretamente, vemos cómo en el grupo denominado Occidental es el que hay más presencia de personas de Barcelona tiene, mientras que en el grupo Enfadado/Frustrado es el que hay menos personas procedentes de esta provincia.

Tabla 15. DIFERENCIAS ESTADÍSTICAMENTE SIGNIFICATIVAS ENTRE CATEGORÍAS DE LAS VARIABLES SOCIODEMOGRÁFICAS Y CADA GRUPO

	Indefinido/ Apático (Negro)	Consciente/ Motivado (Rojo)	Enfadado/ Frustrado (Verde)	Bienestante (Azul)
Nº de individuos	122	92	79	92
%Hombres	34,76%	21,39%	13,37%	30,48%
% Mujeres	28,79%	26,26%	27,27%	17,68%
P-value	0,19377	0,2689	0,0005**	0,0029**
% Barcelona	26,06%	24,30%	22,89%	26,76%
P-value	0,00004	0,70634	0,0506*	0,0248**

Fuente: elaboración propia

5.4. Valores de los ratio de Posición Relativa para España con la ponderación de las medias extraídas de la encuesta para cada variable

Para concluir el trabajo, lo que se ha hecho es la aplicación en España de la ponderación de la a partir de la media aritmética de los cuatro ejes preguntados en la encuesta. Como vimos en el apartado anterior “5.1 Análisis Univariantes”, las medias de los cuatro ejes fueron las siguientes: la del eje de disponibilidad alimentaria de 8,9, la del eje de políticas públicas de 8,5, la del eje de economía de 8,47 y la del eje de saber/factores sociales de 8,37. Teniendo en cuenta que 10 puntos se traduciría en una ponderación de 1, se ha escalado las medias dividiéndola entre 10 y el resultado es una escala de [1,10].

En la siguiente tabla [Tabla 16] vemos, en la parte izquierda los resultados que ya vimos en la fase 2 de la investigación, es decir, la Posición Relativa (PR) de España con un peso igual para cada variable, y en la parte derecha encontramos la PR de cada variable con la ponderación mencionada anteriormente. Los cambios más significativos que hemos observado son:

1. Todas las variables han bajado la PR ya que ninguno de los ejes tenía una puntuación de 10.
2. EL $ITRI_{12}^{inicial}$ ³ es el que se ha visto más afectado, con una evolución negativa del -19,6%, mientras que el $ITRI_{12}^{actual}$ tiene una evolución de -16,4% y el $ITRI_{12}^{futuro}$ de -16,3% .
3. Si nos fijamos en el $ITRI_{12}^{global}$, vemos que tiene una evolución negativa del -16,9%.
4. Como se observó en la etapa 2 de la investigación, se presentaba los valores negativos de

³ Como vimos en el apartado de “Marco teórico”, el ITRI (Índice Triptolemos) indica el equilibrio del sistema alimentario de manera global (económico, disponibilidad alimentaria, político y social) que tiene aquel país. Permite entonces entender las correlaciones y comparaciones entre países y sociedades.

los PR como indicador de cuánto bien está el sistema alimentario. Vemos que la suma de los valores negativos ha tenido una evolución de -14,8%, mientras que los positivos ha sido del -16,1%.

Tabla 16. VALORES DE LOS RATIO DE POSICIÓN RP (POR CADA TIEMPO) PARA ESPAÑA Y CADA VARIABLE CON Y SIN PONDERACIÓN

Sin ponderación					Con ponderación				
España	Inicial	Actual	Futuro	Media	España	Inicial	Actual	Futuro	Media
D1	-0,085	0,041	0,048	0,002	D1	-0,076	0,036	0,043	0,002
D2	-0,122	-0,019	-0,022	-0,055	D2	-0,109	-0,017	-0,020	-0,049
D3	-0,086	-0,08	-0,075	-0,08	D3	-0,077	-0,071	-0,067	-0,071
E1	-0,066	0,155	0,233	0,107	E1	-0,056	0,131	0,197	0,090
E2	-0,074	-0,054	-0,059	-0,062	E2	-0,063	-0,046	-0,050	-0,052
E3	0,054	0,155	0,263	0,157	E3	0,046	0,131	0,222	0,133
P1	0,015	-0,007	-0,03	-0,008	P1	0,013	-0,006	-0,025	-0,007
P2	0,005	-0,096	-0,182	-0,091	P2	0,004	-0,080	-0,152	-0,076
P3	-0,299	-0,208	-0,138	-0,215	P3	-0,250	-0,174	-0,116	-0,180
S1	0,315	0,258	0,206	0,26	S1	0,264	0,216	0,172	0,218
S2	0,481	0,55	0,591	0,541	S2	0,403	0,460	0,495	0,453
S3	0,296	0,217	0,16	0,225	S3	0,248	0,182	0,134	0,188
ITRI ₁₂	0,036	0,076	0,083	0,065	ITRI ₁₂	0,029	0,064	0,069	0,054

Fuente: Elaboración propia

6. Discusión

Este apartado tiene como objetivo comparar, contrastar y discutir los resultados y los procedimientos realizados en el trabajo, así como reconocer las debilidades y fortalezas del estudio.

Uno de los puntos a destacar del trabajo es la capacidad que tiene el modelo planteado en las dos fases de la investigación para describir e interpretar el sistema alimentario global. Se trata de un estudio novedoso ya que por una lado, tiene en cuenta muchos aspectos que influyen en el sistema alimentario y que por otro lado, esto permite tener una prospectiva de 1/3 de la temporalidad introducida en los datos. Lo demuestra el hecho de que sea Cátedra de la UNESCO y que sea de interés para el Estado. Por otro lado, la aplicación del modelo en diferentes países, en concreto de 4 países mediterráneos (España, Francia, Grecia e Italia) le da un valor añadido demostrando la practicidad de la investigación y de observar cómo se ajusta la predicción a la realidad una vez hayan pasado los años estimados.

Por otro lado, introducir unas ponderaciones para aplicar a las variables en el cálculo de indicadores de centralidad permite afinar y ajustar mejor el modelo ya que le da la importancia

específica a cada eje según la valoración de la sociedad en cuestión. Al hacerlo se tiene que tener en cuenta que se trata de una valoración de la ciudadanía. Consecuentemente se le cede el poder a la población y, en contra, la pierde el investigador a la hora de definir la importancia de los ejes. Es decir, ¿se le tiene que dar menos importancia a las políticas públicas porque la sociedad así lo piensa? ¿Estaríamos así potenciando una perspectiva y unos resultados que dependería de la ciudadanía? Es una cuestión que se tiene que tener en cuenta a la hora de presentar los resultados y las interpretaciones.

Entonces, a partir de la técnica cuantitativa de la encuesta se ha podido llevar a cabo el objetivo del trabajo: obtener la valoración de los cuatro ejes de la alimentación de Cataluña de forma representativa ya que se han alcanzado el número necesario de encuestados. Concretamente, sobre este aspecto de la representatividad cabe mencionar ciertas fortalezas y debilidades. Por una lado, una de las fortalezas es la aplicación de estratos para que la muestra, a parte de que sea representativa en global, también lo sea plasmando la diversidad de la sociedad catalana. Específicamente, aplicar tres variables de afijaciones le da un potencial importante en este término. De hecho lo muestra, la variable “clase social”, donde vemos que los resultados se asemejan fuertemente en cómo se compone el universo según el INE y, añadiendo la técnica de ponderaciones de los estratos en función de los encuestados obtenidos, el resultado muestra que refleja de una forma rigurosa Cataluña. Por otro lado, en las debilidades es necesario mencionar dos: una, que no todos los estratos han sido alcanzados. Encontramos por ejemplo que en Lleida tenemos el 66,7% de los encuestados necesarios para que reflejar la diversidad de la provincia. Y dos, que con los resultados de la variable “estudios” vemos un sesgo importante: un 75,8% dice que tiene estudios superiores, mientras que en el INE los datos son del 32,4%. Una de las causas puede deberse a la forma de difusión del cuestionario que, como fuente principal, fue la difusión de los círculos sociales de los investigadores. De manera que estos dos factores ponen en relieve el grado de representatividad de la encuesta.

El análisis bivalente ha servido para conseguir el objetivo de hacer un profiling de los cuatro ejes y las características sociodemográficas que lo determinan. Los resultados muestran lo siguiente:

En el caso de la disponibilidad alimentaria, es decir, el acceso a alimentos nutricionalmente adecuados para la dieta, encontramos que es el eje que se le da una mayor importancia y que concretamente hay una discriminación entre ser mujer o hombre. Sobre el primer resultado decir que no es de extrañar ya que la disponibilidad alimentaria es el primer escalón de toda la cadena.

Si uno no tiene disponibilidad de acceso a alimentos, por mucho que haya economía, buenas políticas y alto grado de saber sobre el tema, la sociedad no se podrá alimentar. Y referente al segundo resultado, una mayor importancia en las mujeres que en los hombres, podría relacionarse con el papel que tiene la mujer dentro del núcleo familiar basado en el de proveer alimentos y en consecuencia, cocinar para sus hijos. De hecho, las mujeres dedican más tiempo a cocinar que los hombres, concretamente 44 minutos más (CIS, 2010).

El eje economía, entendido como tener el dinero necesario para la compra de alimentos, resulta estar en el penúltimo lugar en la valoración de importancia y a la vez no presenta discriminaciones en las variables sociodemográficas. Es decir, la economía aunque pueda parecer algo esencial al vivir en una sociedad capitalista donde por esencia el dinero se entiende como herramienta para la adquisición de “todo”, no está en los ejes más importantes y, no lo es por toda la sociedad catalana, indistintamente del género, la edad y la provincia.

El eje político vemos que es el segundo que se le da más importancia con un 8,5. Y al fijarnos con las características sociodemográficas que lo determinan vemos que es el sexo y la provincia. Concretamente, en referencia al sexo vemos que las mujeres le dan más importancia que los hombres. Es un resultado peculiar y difícil de interpretar desde el punto de vista social ya que por ejemplo, en la política encontramos mucha más presencia de hombres que de mujeres. Por tanto, se trata de un hilo del cual sería interesante investigar en más profundidad. Por otro lado, en el caso de la provincia se observa como en Barcelona se le da menos importancia que al resto. Una de las causas podría venir definida por el descontento político más enfocada en la área metropolitana que en el resto aunque no hemos encontrado ningún informe que lo pudiera corroborar.

Y finalmente el eje de aspectos culturales es el que se le ha dado una menor importancia con un 8,37 puntos. Y al observar el profiling vemos diferencia en las edades: el rango de más edad (de 45 o más) muestra más interés en los aspectos culturales que los menores de 45 años. En este aspecto hay que prestar atención al cambio cultural y el estilo de vida que se produjo en los años 60 en la sociedad española y que, consecuentemente, tuvo repercusiones en la alimentación. La sociedad del consumo y el cambio en la familia debido a que las dos figuras parentales trabajasen, el tiempo se redujo considerablemente y entro en juego la comida rápida y la precocinada (Terrejón, R., 2016). Este fenómeno puede explicar el hecho que las personas mayores, las que han vivido la cultura de la alimentación tradicional, le dan más importancia a las que han convivido con

la comida rápida y a la reducción de tiempo dedicado a cocinar.

Por otro lado, con el clúster hemos podido definir los diferentes perfiles según las variables que hacían referencia a la alimentación. El resultado ha sido interesante ya que los grupos están fuertemente discriminados. Encontramos 4 perfiles diferentes: el consciente/motivado, el bienestante, el apático y el frustrado. Hay tres aspectos a destacar del análisis.

El primero es el hecho de que el grupo denominado “bienestante” por un lado, le da importancia a los ejes de economía, de disponibilidad alimentaria y de políticas públicas, por otro lado, lo que discrimina es la categoría “Incrementar la calidad de los alimentos para mejorar su valor nutricional y sabor” de la variable sobre qué papel tiene que desarrollar la ciencia y también que se trata de un grupo donde hay diferencias significativas entre sexo, con más presencia de hombres. Es un resultado interesante y que muestra la contrariedad y el egoísmo de la especie humana como explicaba ya Nietzsche (De los Cobos, A., 1998). Tanto el eje de disponibilidad de acceso de alimentos, como el de economía al vivir en un sistema capitalista como el político al ser una sociedad estructurada a partir de las reglas establecidas por el Estado, son anteriores e imprescindibles para luego poder incrementar la calidad de los alimentos. Aun son muchos los datos que presentan la pobreza que hay en el planeta, de manera que este grupo refleja el egoísmo en el hecho de que den importancia a los ejes necesarios para erradicar la pobreza pero en cambio, prefieran incrementar la calidad en vez de la producción para que haya una mayor disponibilidad de alimentos en la población.

El segundo a destacar es cómo el tipo de alimentación no es un factor discriminante de los diferentes perfiles. De manera que el tipo de alimentación menos sostenible y el que menos favorece a una disponibilidad de acceso de alimentos a toda la sociedad que es el de comer carne, no está correlacionado a que la gente se posicione en mayor medida a que el papel a desarrollar por la ciencia sea el de priorizar el valor nutricional de los alimentos verso a aumentar la producción de alimentos para que haya una mayor disponibilidad en la población.

Y el tercero es el hecho de que el grupo más grande sea el “Apático/Indefinido”, representando el 31,7% de los individuos. Sociólogos como Bauman, a través de conceptos como la sociedad líquida explican esta apatía de la sociedad basado en que el ser humano no se cree capaz y con el poder suficiente para poder hacer algún cambio y en consecuencia se muestra apático y se posiciona al margen de los problemas de la sociedad esperando que ese “alguien” (que sí tienen el poder pero

difícil de identificar) lo haga.

Finalmente, con la aplicación de la ponderación de las valoraciones en el modelo cuantitativo de España, vemos como el $ITRI_{12}^{global}$ tiene una evolución negativa del -16,9% y que cambian las posiciones relativas (PR) de las 12 variables que lo forman. El resultado más relevante es ver cómo cambian la suma de valores negativos y positivos entre el modelo con la ponderación y sin la ponderación. Como vimos en el marco teórico la PR nos indica cuánto de bien está el país: si es menor que 0 significa que tiene una posición por debajo de la media y si es mayor que 0 es que está por encima de la media. De manera que sin ponderación la suma de valores negativos tiene una evolución de -14,8%, mientras que los positivos es de -16,1%. Esto significa que las dimensiones del modelo que iban mejor son las que se les da menos importancia mientras que las dimensiones del modelo de España que tenían una PR por debajo de 0 han tenido una evolución negativa mayor, lo que se traduce a que son las importantes para la sociedad estudiada.

7. Conclusiones

Como se ha podido constatar, este trabajo trata de un tema muy esencial ya que es uno de los pilares indispensables para la supervivencia. Se ha mostrado que es posible la cuantificación matemática de un modelo que englobe muchas de las perspectivas que influyen al sistema alimentario y además con una visión prospectiva. A partir del planteamiento de la encuesta como herramienta para identificar las valoraciones de los 4 ejes: disponibilidad alimentaria, economía, políticas públicas y saber, se ha alcanzado el objetivo del trabajo: ver como cambia el modelo planteado con estas valoraciones y en concreto, el Índice Triptolemos.

Luego con el análisis de los datos se ha podido identificar qué perfiles demográficos determinan cada eje y cómo se dividen los individuos de la sociedad catalana en función a las variables preguntadas sobre el sistema alimentario. Hemos visto como los resultados son muy interesantes y que evidencian problemas sociales actuales sobre el tema o que repercuten en él, lo que le da un potencial elevado para posteriores investigaciones. De hecho este trabajo pretende ser, como apuntan los investigadores principales (Ramón Clotet y Eusebi Jarauta), un proceso estándar y sistematizado para implementarse en posteriores estudios en otras universidades que tienen convenio, con especial interés en las universidades de diferentes continentes para tener la visión global del sistema alimentario (objetivo principal de la investigación). Por tanto, este trabajo

aunque estadísticamente no tenga modelos matemáticos complejos, sí que tiene un fuerte componente de practicidad, lo que conlleva a ser una tesis con gran impacto.

Por otro lado, este trabajo ha conseguido contrastar las hipótesis planteadas:

- Se ha constatado cómo el eje de economía y disponibilidad alimentaria se le da mucha importancia aunque éste último, en un nivel muy similar al de las políticas públicas
- Sí hay diferencias sociodemográficas significativas para cada eje pero que la edad es la menos influyente y finalmente
- Afirmar la hipótesis de que Cataluña como país de tendencia occidental le da una mayor importancia a la soberanía alimentaria (detenimiento al sabor nutricional de los alimentos) que a la globalización alimentaria (mayor disponibilidad de alimentos en la sociedad).

Y finalmente decir que a partir de este trabajo he aprendido mucho sobre el tema de la encuesta, la depuración de los datos, la necesidad de tener variables control para eliminar posibles duplicados, la técnica de estratificación y ponderaciones por afijaciones, la diversidad de técnicas de imputación de valores perdidos, metodologías para mejorar la rigurosidad de los datos como es utilizar la distancia Gower para el clúster ya que permite la variedad de clases de variables,... así como aplicar lo aprendido en el máster como es la técnica de bootstrapping para solucionar la no normalidad, realizar un análisis de clúster para determinar perfiles, los análisis bivariantes que han permitido hacer un profiling de los ejes,... Combinar también lo aprendido en la carrera de sociología como es la definición del cuestionario, la realización de las preguntas de manera que cada categoría de respuesta tenga el mismo peso, aplicar softwares específicos para poder hacer difusión,... Se trata entonces de un trabajo que me ha ayudado a ver todo lo aprendido en el máster y darme cuenta que la combinación de la matemática aplicada, la estadística y la sociología se compaginan muy bien y da resultados muy interesantes, dándome energía para seguir avanzando y aprendiendo en este campo de los datos.

BIBLIOGRAFÍA

- Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(2), 139-160. Recuperado de: <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1982.tb01195.x>
- Anguita, J. C., Labrador, J. R., Campos, J. D., Casas Anguita, J., Repullo Labrador, J., & Donado Campos, J. (2003). La encuesta como técnica de investigación. Elaboración de cuestionarios y tratamiento estadístico de los datos (I). *Atención primaria*, 31(8), 527-538. Recuperado de <https://www.survenia.com/articles/articulo1.pdf>
- Azur, M. J., Stuart, E. A., Frangakis, C., & Leaf, P. J. (2011). Multiple imputation by chained equations: what is it and how does it work?. *International journal of methods in psychiatric research*, 20(1), 40-49. doi:10.1002/mpr.329. Recuperado de: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3074241/>
- Azzalini, A., & Menardi, G. (2013). Clustering via nonparametric density estimation: The R package pdfCluster. *arXiv preprint arXiv:1301.6559*. Recuperado de: <https://arxiv.org/pdf/1301.6559.pdf>
- Badii, M. H., Castillo, A. W., & Landeros, J. (2017). Precisión de los índices estadísticas: Técnicas de jackknife & bootstrap. *Innovaciones de negocios*, 4(7). Recuperado de: https://www.researchgate.net/publication/315812679_Precision_de_los_indices_estadisticas_Tecnicas_de_jackknife_bootstrap_Precision_of_statistical_indices_Jackknife_bootstrap_techniques
- Bauman, Z. (2015). *Modernidad líquida*. Fondo de cultura económica.
- Clotet, R., Colomer, Y., Jarauta, E., & Mayor, F. (2013). *El sistema alimentario global: I-Definición de un espacio* (No. 1102-2016-91119, pp. 13-32). Recuperado de: http://www.triptolemos.org/wp-content/uploads/2016/07/MAGRAMA_SAG.pdf
- Colom, S. M., Ajenjo, M., & Català, V. B. (2018). La masculinización del tiempo dedicado al trabajo doméstico rutinario. *Reis: Revista española de investigaciones sociológicas*, (163), 41-58. Recuperado de: http://www.reis.cis.es/REIS/PDF/REIS%20_163_031522757991605.pdf
- De los Cobos, A., (1998). *Carácter Social y Enfermedad de la metáfora y las ciencias sociales*. Universidad de Castilla-La Mancha, 40. Recuperado de: https://books.google.es/books?id=bj5J_q7pJkgC&pg=PA40&dq=egoismo+del+ser+humano+nietzsche&hl=ca&sa=X&ved=0ahUKewiPwr2A0fXkAhWilFwKHW7bCQUQ6AEINDAB#v=onepage&q&f=false
- Echauri, A. M. F., Minami, H., & Sandoval, M. J. I. (2014). La Escala de Likert en la evaluación docente: acercamiento a sus características y principios metodológicos. *Perspectivas docentes*, (50). Recuperado de: <https://dialnet.unirioja.es/servlet/articulo?codigo=6349269>
- Filaire, T. (2018). Towards Data Science: Clustering on mixed type data. URL: <https://towardsdatascience.com/clustering-on-mixed-type-data-8bbd0a2569c3>
- Gómez, J. L. P. (2002). Estrategias de ponderación de la respuesta en encuestas de satisfacción de usuarios de servicios. *Metodología de encuestas*, 4(2), 175-193. Recuperado de: <http://casus.usal.es/pkp/index.php/MdE/article/view/923>

Instituto Nacional de Estadística (INE). (2019). Población de 16 y más años por nivel de formación alcanzado, sexo y comunidad autónoma. Porcentajes respecto del total de cada comunidad. *Población en viviendas familiares*. Recuperado de: <http://www.ine.es/jaxiT3/Tabla.htm?t=6369>

Jarauta, E., Colomer, Y., & Clotet, R. (2018). El sistema alimentario global: II aproximación cuantitativa al espacio agroalimentario de la Europa mediterránea. *Revista española de estudios agrosociales y pesqueros*, (249), 15-38. Recuperado de: https://www.mapa.gob.es/ministerio/pags/Biblioteca/Revistas/pdf_REEAP_%2FPdf_REEAP_r249_15_38.pdf

Jhon. Ggplot2 pie chart : Quick start guide - R software and data visualization. Australia: Sthda. Recuperado de: <http://www.sthda.com/english/wiki/ggplot2-pie-chart-quick-start-guide-r-software-and-data-visualization#infos>

Maechler, M. (2019). "Finding Groups in Data": Cluster Analysis Extended Rousseeuw et al. Package "Cluster". CRAN. Recuperado de: <https://cran.r-project.org/web/packages/cluster/cluster.pdf>

Matthias, S. (2019). Clustering of Weighted Data: Package 'weightedCluster'. *Cran.r-project*. Recuperado de: <https://cran.r-project.org/web/packages/WeightedCluster/WeightedCluster.pdf>

Monleón-Getino, T. (2016). Novedades en el ANOVA Diseño de experimentos, su análisis y diagnóstico. *Group of Research on Statistics and Bioinformatics (GRBIO): University of Barcelona*. Recuperado de: https://www.researchgate.net/profile/Antonio_Monleon-Getino/publication/304283596_Disenio_de_experimentos_su_analisis_y_diagnostico/links/576b8cea08aefcf135bd5977.pdf

Pasek, J. (2018). Weighting and Weighted Statistics: Package 'weights'. *Cran.r-project*. Recuperado de: <https://cran.r-project.org/web/packages/weights/weights.pdf>

Ruiz, J. F. M. (2005). Los factores definitorios de los grandes grupos de edad de la población: tipos, subgrupos y umbrales. *Scripta Nova: revista electrónica de geografía y ciencias sociales*, 9(2). Recuperado de: <http://www.ub.edu/geocrit/sn/sn-190.htm>

Torrejón, R., (2016). El auge de la comida rápida en España. *Variación21: El periodismo universitario en Internet*. Recuperado de: <http://variacionxxi.com/2016/01/13/franquicias-comida-rapida/>

Triptolemos. (2019). Cátedra UNESCO Science and Innovation for Sustainable Development: Global Food Production and Food Safety. *Proyecots: Cétedra Unesco*. Recuperado de: <http://www.triptolemos.org/catedra-unesco/>

Triptolemos. (2019). El ministro de agricultura conoce de primera mano la labor de la fundación triptolemos. *Noticias*. Recuperado de: <http://www.triptolemos.org/luis-planas-conoce-primera-mano-la-labor-la-fundacion-triptolemos/>

ANEXO I Cuestionario

Formato online:

<https://forms.gle/955ZzgWLGSLvfUdB8>

Formato papel:

Encuesta sobre el sistema alimentario. 2019



Ha accedido al formulario relativo a una encuesta que se lleva a cabo en el marco de una investigación universitaria de la UPC (<https://futur.upc.edu/buscar/t/amFyYXV0YQ==>), con colaboración de la Fundación Triptolemos (<http://www.triptolemos.org/>, <http://www.sistemaalimentarioglobal.org>).

Esta investigación tiene como objetivo profundizar en el estudio sobre el sistema alimentario global que promueve la Fundación Triptolemos y, principalmente, determinar la ponderación de cuatro factores relacionados con el sistema alimentario.

A continuación encontrará dos apartados: el primero sobre datos sociodemográficos y el otro sobre la ponderación de los cuatro factores de la alimentación.

Ante cualquier duda se puede poner en contacto con nosotros a través de francisca.morey@estudiant.upc.edu

DATOS SOCIODEMOGRÁFICOS

1. Sexo

Marque sólo una opción

☐

Mujer

☐

Hombre

2. Fecha de nacimiento

Especifique el día, mes y año de nacimiento

--	--

Día

--	--

Mes

--	--	--	--

Año

3. Comarca de residencia

.....

4. ¿Cuál es su nivel de estudios?

Marque con una cruz sólo una opción

☐

Educación infantil (0 a 6 años)

- ☐ Educación primaria
- ☐ Educación secundaria (ESO)
- ☐ Educación postsecundaria (Bachillerato y grados formativos)
- ☐ Educación universitaria (grado, diplomaturas, máster, doctorado,...)
- ☐ Educación no formal

5. Si tuviera que elegir una clase social, ¿a cuál diría que pertenece?

Marque con una cruz sólo una opción

- ☐ Baja
- ☐ Media-baja
- ☐ Media
- ☐ Media-alta
- ☐ Alta

6. En que tipo de alimentación se encuentra más cómodo:

Marque con una cruz sólo una opción

- ☐ Predominancia de vegetales
- ☐ Sin preferencias (comer de todo)
- ☐ Predominancia de carne y derivados

VALORACIÓN DE LA IMPORTANCIA DE CUATRO FACTORES ALIMENTARIOS

En la sociedad actual el poder disponer de una alimentación adecuada es un resultado de muchos factores. Por ello, nos gustaría saber su opinión con criterios de ética y sostenibilidad, sobre la importancia de cuatro de ellos. Responda pensando en donde vive (Cataluña):

Marque con una cruz sólo una opción

7. Tener la disponibilidad de acceso a alimentos nutricionalmente adecuados a la dieta:

	1	2	3	4	5	6	7	8	9	10	
Poco importante	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Muy importante

8. Tener el nivel económico para comprar los alimentos necesarios:

	1	2	3	4	5	6	7	8	9	10	
Poco importante	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Muy importante

9. Las políticas públicas para favorecer la producción de alimentos y la seguridad alimentaria:

	1	2	3	4	5	6	7	8	9	10	
Poco importante	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Muy importante

10. Los factores culturales, hábitos alimentarios, la formación y los placeres gastronómicos:

	1	2	3	4	5	6	7	8	9	10	
Poco importante	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Muy importante

11. En relación con los siguientes dos retos a desarrollar por la ciencia, ¿cuál cree Ud. que es el más importante?

- ☐ Incrementar la producción de alimentos para una mayor disponibilidad en la población
- ☐ Incrementar la calidad de los alimentos para mejorar su valor nutricional y sabor

Gracias por participar!

Los datos serán tratados de manera totalmente confidencial, analizados estadísticamente y utilizados con el fin de realizar y / o apoyar el proyecto "El sistema alimentario global". Muchas gracias por participar!

ANEXO II CÓDIGO R

```
library(readxl)
library(base)
library(mice)
library(varComp)
library(data.table)
library(sqldf)
library(data.table)
library(ggplot2)
library(dplyr)
library(sqldf)
library(weights)
library(haven)
```

```

library(factoextra)
library(StatMatch)
library(cluster)
library(MVN)
library(Hotelling)

#####IMPORTACIÓ DE LOS DATOS#####
df <- read_excel("E:/TFM/BASE DE DADES1.xls")
df <- as.data.table(df)

#####IMPUTACIÓ DATOS PERDIDOS#####
na <- data.frame(md.pattern(TFM))
input <- df
input$Any.de.naixement <- as.numeric(input$Any.de.naixement)
input$imputats[!input$any == ""] <- 1
set.seed(345)
tempData <- mice(input, meth = "norm.boot")
completedData <- complete(tempData, 5)
completedData$any <- round(completedData$any, 0)

#####RECODIFICACIÓ Y AGRUPACIÓ DE VARIABLES#####
df$Edatdrup <- cut(df$edat, c(0, 24, 34, 44, 100), dig.lab = 2)

df[edat <= 24, Edad_2 := 'De 16 a 24 anos']
df[edat > 24 & edat <= 34, Edad_2 := 'De 25 a 34 anos']
df[edat > 34 & edat <= 44, Edad_2 := 'De 34 a 44 anos']
df[edat > 45, Edad_2 := 'De 45 o mas anos']

df[Edad_2 == 'De 16 a 24 anos' | Edad_2 == 'De 25 a 34 anos' | Edad_2 == 'De 34 a 44 anos', Edad_3 := 0 ]
df[Edad_2 == 'De 45 o mas anos', Edad_3 := 1 ]

df[Sexe == 'Home', CodSexo := 1]
df[Sexe == 'Dona', CodSexo := 2]

df[Provincia == 'Barcelona', CodProvincia := 1]
df[Provincia == 'Girona', CodProvincia := 2]
df[Provincia == 'Lleida', CodProvincia := 3]
df[Provincia == 'Tarragona', CodProvincia := 4]

df[Visio_futur == 'Incrementar la qualitat dels aliments per millorar el seu valor nutricional i sabor', CodVisio_futur := 0 ]
df[Visio_futur == 'Incrementar la producció d'aliments per una major disponibilitat en la població', CodVisio_futur := 1 ]

df[Tipus_alim == 'Sense preferències (menjar de tot)', CodTipus_alim := 1 ]
df[Tipus_alim == 'Preponderància de vegetals', CodTipus_alim := 2 ]
df[Tipus_alim == 'Preponderància de carn i derivats', CodTipus_alim := 3 ]

df[Clase == 'Baixa', CodClase := 1 ]
df[Clase == 'Mitjana-baixa', CodClase := 2 ]
df[Clase == 'Mitjana', CodClase := 3 ]
df[Clase == 'Mitjana-alta', CodClase := 4 ]
df[Clase == 'Alta', CodClase := 5 ]

df[Estudis == 'Educació infantil (0 a 6 anys)', CodEstudis := 1 ]
df[Estudis == 'Educació primària', CodEstudis := 2 ]
df[Estudis == 'Educació secundària (ESO)', CodEstudis := 3 ]
df[Estudis == 'Educació postsecundària (Batxillerat i graus formatius)', CodEstudis := 4 ]

```

```

df[Estudis=='Educació universitària (grau, diplomatures, màster, doctorat,...)', CodEstudis:= 5 ]

df$Provincia <- as.factor(df$Provincia)
df <- cbind(df, data.table(dummify(df$Provincia)))

#####PONDERACIONES#####
df[CodSexo==1 & CodProvincia==1 & Edad_3==0, Ponderacion_2:= 0.890]
df[CodSexo==1 & CodProvincia==1 & Edad_3==1, Ponderacion_2:=1.220]

df[CodSexo==1 & CodProvincia==2 & Edad_3==0, Ponderacion_2:= 1.500]
df[CodSexo==1 & CodProvincia==2 & Edad_3==1, Ponderacion_2:=0.909]

df[CodSexo==1 & CodProvincia==3 & Edad_3==0, Ponderacion_2:= 1]
df[CodSexo==1 & CodProvincia==3 & Edad_3==1, Ponderacion_2:= 2.0]

df[CodSexo==1 & CodProvincia==4 & Edad_3==0, Ponderacion_2:= 1.125]
df[CodSexo==1 & CodProvincia==4 & Edad_3==1, Ponderacion_2:=1.833]

df[CodSexo==2 & CodProvincia==1 & Edad_3==0, Ponderacion_2:= 0.543]
df[CodSexo==2 & CodProvincia==1 & Edad_3==1, Ponderacion_2:=0.988]

df[CodSexo==2 & CodProvincia==2 & Edad_3==0, Ponderacion_2:= 0.6]
df[CodSexo==2 & CodProvincia==2 & Edad_3==1, Ponderacion_2:=1]

df[CodSexo==2 & CodProvincia==3 & Edad_3==0, Ponderacion_2:= 0.8]
df[CodSexo==2 & CodProvincia==3 & Edad_3==1, Ponderacion_2:=6]

df[CodSexo==2 & CodProvincia==4 & Edad_3==0, Ponderacion_2:= 1]
df[CodSexo==2 & CodProvincia==4 & Edad_3==1, Ponderacion_2:= 2.0]

#####ANÁLISIS UNIVARIANTE#####
wtd.hist(df$Disponibilitat, freq = FALSE, weight=df$Ponderacion_2, main = "Tener la disponibilidad de acceso a
alimentos
    nutricionalmente adecuados a la dieta:",
    xlab = "Valoración", ylab = "Densidad", plot = TRUE, labels =TRUE)
wtd.hist(df$Economia, freq = FALSE, weight=df$Ponderacion_2, main = "Tener el nivel económico para comprar los
alimentos necesarios:",
    xlab = "Valoración", ylab = "Densidad", plot = TRUE, labels =TRUE)
wtd.hist(df$Politiques, freq = FALSE, weight=df$Ponderacion_2, main = "Las políticas públicas para favorecer la
producción
    de alimentos y la seguridad alimentaria:",
    xlab = "Valoración", ylab = "Densidad", plot = TRUE, labels =TRUE)
wtd.hist(df$Saber, freq = FALSE, weight=df$Ponderacion_2, main = "Los factores culturales, hábitos alimentarios, la
formación y
    los placeres gastronómicos:",
    xlab = "Valoración", ylab = "Densidad", plot = TRUE, labels =TRUE)
wtd.hist(df$CodTipus_alim, freq = FALSE, weight=df$Ponderacion_2, main = "Tipo de alimentación",
    xlab = "Valoración", ylab = "Densidad", plot = TRUE, labels =TRUE)
wtd.hist(df$CodVisio_futur, freq = FALSE, weight=df$Ponderacion_2, main = "Visió futur",
    xlab = "Valoración", ylab = "Densidad", plot = TRUE, labels =TRUE)
wtd.hist(df$CodClase, freq = FALSE, weight=df$Ponderacion_2, main = "Clase Social",
    xlab = "Valoración", ylab = "Densidad", plot = TRUE, labels =TRUE)
wtd.hist(df$CodEstudis, freq = FALSE, weight=df$Ponderacion_2, main = "Estudios",
    xlab = "Valoración", ylab = "Densidad", plot = TRUE, labels =TRUE)

#####ANÁLISIS BIVARIANTE#####
####SEXO####

```

```

#Disponibilidad:
ysplit <- split(df$Disponibilitat, df$CodSexo)
wsplit <- split(df$Ponderacion_2, df$CodSexo)
wtd.t.test(ysplit[[1]], ysplit[[2]], wsplit[[1]], wsplit[[2]], bootse = TRUE, bootp=TRUE, bootn = 50000)

#Economía:
ysplit <- split(df$Economia, df$CodSexo)
wsplit <- split(df$Ponderacion_2, df$CodSexo)
wtd.t.test(ysplit[[1]], ysplit[[2]], wsplit[[1]], wsplit[[2]], bootse = TRUE, bootp=TRUE, bootn = 50000)

#Polítiques:
ysplit <- split(df$Politiques, df$CodSexo)
wsplit <- split(df$Ponderacion_2, df$CodSexo)
wtd.t.test(ysplit[[1]], ysplit[[2]], wsplit[[1]], wsplit[[2]], bootse = TRUE, bootp=TRUE, bootn = 50000)

#Saber:
ysplit <- split(df$Saber, df$CodSexo)
wsplit <- split(df$Ponderacion_2, df$CodSexo)
wtd.t.test(ysplit[[1]], ysplit[[2]], wsplit[[1]], wsplit[[2]], bootse = TRUE, bootp=TRUE, bootn = 50000)

####Por provincia#####
####Disponibilidad
wtd.chi.sq(df$Disponibilitat,df$CodProvincia, weight = df$Ponderacion_2)

####Economía
wtd.chi.sq(df$Economia,df$CodProvincia, weight = df$Ponderacion_2)

####Políticas
wtd.chi.sq(df$Politiques,df$CodProvincia, weight = df$Ponderacion_2)
#Barcelona
ysplit <- split(df$Politiques, df$Barcelona)
wsplit <- split(df$Ponderacion_2, df$Barcelona)
wtd.t.test(ysplit[[1]], ysplit[[2]], wsplit[[1]], wsplit[[2]], bootse = TRUE, bootp=TRUE, bootn = 50000)

#Girona
ysplit <- split(df$Politiques, df$Girona)
wsplit <- split(df$Ponderacion_2, df$Girona)
wtd.t.test(ysplit[[1]], ysplit[[2]], wsplit[[1]], wsplit[[2]], bootse = TRUE, bootp=TRUE, bootn = 50000)

#Tarragona
ysplit <- split(df$Politiques, df$Tarragona)
wsplit <- split(df$Ponderacion_2, df$Tarragona)
wtd.t.test(ysplit[[1]], ysplit[[2]], wsplit[[1]], wsplit[[2]], bootse = TRUE, bootp=TRUE, bootn = 50000)

#Lleida
ysplit <- split(df$Politiques, df$Lleida)
wsplit <- split(df$Ponderacion_2, df$Lleida)
wtd.t.test(ysplit[[1]], ysplit[[2]], wsplit[[1]], wsplit[[2]], bootse = TRUE, bootp=TRUE, bootn = 50000)

###Saber
wtd.chi.sq(df$Saber,df$CodProvincia, weight = df$Ponderacion_2)

####Edad
ysplit <- split(df$Disponibilitat, df$Edad_3)
wsplit <- split(df$Ponderacion, df$Edad_3)
wtd.t.test(ysplit[[1]], ysplit[[2]], wsplit[[1]], wsplit[[2]], bootse = TRUE, bootp=TRUE, bootn = 50000)

```

```

ysplit <- split(df$Economia, df$Edad_3)
wsplit <- split(df$Ponderacion_2, df$Edad_3)
wtd.t.test(ysplit[[1]], ysplit[[2]], wsplit[[1]], wsplit[[2]], bootse = TRUE, bootp=TRUE, bootn = 50000)

ysplit <- split(df$Politiques, df$Edad_3)
wsplit <- split(df$Ponderacion_2, df$Edad_3)
wtd.t.test(ysplit[[1]], ysplit[[2]], wsplit[[1]], wsplit[[2]], bootse = TRUE, bootp=TRUE, bootn = 50000)

ysplit <- split(df$Saber, df$Edad_3)
wsplit <- split(df$Ponderacion_2, df$Edad_3)
wtd.t.test(ysplit[[1]], ysplit[[2]], wsplit[[1]], wsplit[[2]], bootse = TRUE, bootp=TRUE, bootn = 50000)

#####ANÁLISIS DE CLÚSTER#####
#Matriz de distancias:
dai.b <- daisy(df, metric = 'gower', type = list(ordratio=1:4, asymm = 5:8), weights = c(0.89, 0.845,
0.837,0.837,1,0.333,0.333,0.333))# me falta posar les ponderacions

#Dendograma
cluster.G.W <- hclust(dai.b, method="ward.D2")
plot(cluster.G.W ,hang = -1, cex=0.7)

#Test Silhouette
sil<-fviz_nbclust(dfnum, pam, diss=dist(dfnum), method = "silhouette", nboot = 500000)
sil

#Clúster no jerárquico
k<-4
pam.k <- pam(dai.b,k,diss=T)
pam.k$medoids

#Escalamiento multidimensional
mds.pam <- cmdscale(d.hel, eig=TRUE)
plot(mds.pam$points[,1], mds.pam$points[,2], main="PAM", xlab="Axis 1", ylab="Axis 2",
col=pam.k$cluster, pch=19)
text(mds.pam$points[,1], mds.pam$points[,2], labels= NULL, pos=1, cex=0.5, offset=0.15)

Axis1<-round(cor(mds.pam$points[,1], df),3)
Axis1
Axis2<-round(cor(mds.pam$points[,2], df),3)
Axis2

#Profiling del clúster:
Grupo$cluster <- c(pam.k$clustering)

Grupo1 <- data.table(sqldf("select * from Grupo where cluster is 1"))
Grupo2 <- data.table(sqldf("select * from Grupo where cluster is 2"))
Grupo3 <- data.table(sqldf("select * from Grupo where cluster is 3"))
Grupo4 <- data.table(sqldf("select * from Grupo where cluster is 4"))

wtd.chi.sq(df$cluster,df$CodSexo, weight = df$Ponderacion_2)
wtd.chi.sq(df$cluster,df$Edad_3, weight = df$Ponderacion_2)
wtd.chi.sq(df$cluster,df$CodProvincia, weight = df$Ponderacion_2)

#Diferencias significativas entre grupos de la variable sexo
wtd.hist(Grupo1$CodSexo, weight=Grupo1$Ponderacion_2, main = "Grupo1$CodSexo",
xlab = "Valoración", ylab = "Densidad", plot = TRUE, labels =TRUE)
wtd.hist(Grupo2$CodSexo, weight=Grupo2$Ponderacion_2, main = "Grupo2$CodSexo",

```

```

      xlab = "Valoración", ylab = "Densidad", plot = TRUE, labels =TRUE)
wtd.hist(Grupo3$CodSexo, weight=Grupo3$Ponderacion_2, main = "Grupo3$CodSexo",
      xlab = "Valoración", ylab = "Densidad", plot = TRUE, labels =TRUE)
wtd.hist(Grupo4$CodSexo, weight=Grupo4$Ponderacion_2, main = "Grupo4$CodSexo",
      xlab = "Valoración", ylab = "Densidad", plot = TRUE, labels =TRUE)

#Diferencias significativas entre grupos de la variable Provincia
wtd.hist(Grupo1$CodProvincia, weight=Grupo1$Ponderacion_2, main = "Grupo1$CodProvincia",
      xlab = "Valoración", ylab = "Densidad", plot = TRUE, labels =TRUE)
wtd.hist(Grupo2$CodProvincia, weight=Grupo2$Ponderacion_2, main = "Grupo2$CodProvincia",
      xlab = "Valoración", ylab = "Densidad", plot = TRUE, labels =TRUE)
wtd.hist(Grupo3$CodProvincia, weight=Grupo3$Ponderacion_2, main = "Grupo3$CodProvincia",
      xlab = "Valoración", ylab = "Densidad", plot = TRUE, labels =TRUE)
wtd.hist(Grupo4$CodProvincia, weight=Grupo4$Ponderacion_2, main = "Grupo4$CodProvincia",
      xlab = "Valoración", ylab = "Densidad", plot = TRUE, labels =TRUE)

```